

13	1322	5.1	474.2	72	AE015297
14	1322.2	5.1	1396.5	72	AE015298
c 15	101.8	4.2	2119.1	72	H01313.194
c 16	101	4.2	6192	5	AR030811
c 17	101	4.2	644.9	5	AR030812
c 18	100.6	4.2	1593.1	72	HHV66AGNM
c 19	97.8	4.1	3296.1	47	AC026685
c 20	96.2	4.0	1924.3	45	AC023920
c 21	95	3.9	1678.0	43	AC020678
c 22	94	3.9	2276.5	49	AC044845
c 23	93	3.9	1177	32	PCSEPT
c 24	88.4	3.7	1463.8	91	HSBA119.119
c 25	86.8	3.6	381.39	92	H0MPANMT
c 26	86.8	3.6	1621.25	43	AC0021594
c 27	83.2	3.5	1336.61	73	U93872
c 28	83	3.4	1336.61	73	U93872
c 29	82.2	3.4	359.99	66	AC0078925
c 30	82.2	3.4	1636.24	52	AC0073862
c 31	82.2	3.4	2528.87	52	AC0073527
c 32	81	3.4	2181.4	31	CEH95E23
c 33	79.8	3.3	2483	11	AF112453
c 34	79.4	3.3	2370	12	MAU056918
c 35	79.2	3.3	1326	32	PMU09766
c 36	79.2	3.3	1326	32	S690114
c 37	79.2	3.3	2187.35	39	AC011899
c 38	78.8	3.3	2751	11	AF116947
c 39	78.8	3.3	3038.19	30	AE003509
c 40	76.6	3.2	3941	12	RNLAR2
c 41	76	3.2	1545	32	PPACSG6
c 42	76	3.2	7218	5	I66494
c 43	75.8	3.1	1072	73	HS21EA
c 44	75.8	3.1	2423	71	AF109134
c 45	75.8	3.1	6670	11	AF112980
					AF015297 Human hepe
					AF015298 Human hepe
					U13194 Human hepe
					AR030811 Sequence
					AR030812 Sequence
					H83431 Human hepe
					AC026285 Homo sapi
					AC023920 Homo sapi
					AC020678 Homo sapi
					AC044845 Mus muscu
					X74358 P. carnea po
					AL117339 Human DNA
					J05582 Human pancr
					AC000040 Homo septi
					AC021554 Homo sapi
					U93872 Kaposi's sa
					AC0078925 Homo sapi
					AC0073862 Homo sapi
					AC0073527 Homo sapi
					Z96102 Caenorhabdi
					AF117453 Homo sapi
					U36519 Mesocricetu
					U09766 Mesocricetu
					S690114 Plasmodium
					AC011899 Homo sapi
					AF116947 Macaca mu
					AE003509 Drosophila
					H83546 R. norvegicu
					J03992 Plasmodium
					I66494 Sequence 14
					M29385 Herpes simp
					AF109134 Homo sapi
					AF112980 Homo sapi

/protein_id="AA064406.1"
 /db_xref="GI:7595307".
 translation="NDPDCDSTWEDEDEDAEDADEDCEDGAGAGARDADGDDEH
 SEBRAPRRSSFSQSMRTGSRNMRATRDIMCYRHRNYPDLYVERONDPDTPSYREIIT
 FLPCNCFIEDILDMWTNDNYLDLENNHSYIQMWFLAREPGVNNHAPLTLIREYFKSS
 ODIQILVRAYAEIMLGYFRIETREDRGYTGKGRANQYKRFQNLNHRSHNLELITRIL
 SCIESLSEHPQAPLVYRFFLEETLYRRELPEYRQSLADYIMFVNRCHORQOLYHFAEM
 HRPKCFEVMGPODKKLRRFPKSSLPFLIESRQSODGAGHEDPEPLISESKKKLLSTR
 SKGGKRVDEGQPSRVSVEPDQAGFLERSQDEGAGHEDPEPLISESKKKLLSTR
 KQPLPPEGQPSASEVEKIALNLEGALCSQSGYRTGQEOEGDQPEANQVQRFQYLR
 VADKRRKRRKYDECTGSAAVVAGGAQKPTIALAGSPASGHRPAGSGENSEVEDEEGAT
 GKKESTPGSPSETPGSPGAPAGVAGGAQKPTSETPGSPGAPTRDEPASPSETPGSP
 AGPADDEAPESSETPGPRAGAGADDEPKITSETPGSPGAPTRDEPASPSETPGSP
 RPAGAGADDEAPESSETPGPRPAGGADDEPASPSETPGSPGAPTRDEPAPKAGEAL
 LQDAEVSSASNGCP"
 BASE COUNT 518 a 778 c 861 g 326 t
 ORIGIN

[illegible]

[illegible]

QY	1861	acaagggatgagccagccgagagcccatctggaaccccccaagcccccgaggaacct	1920
Db	1861	ACAAGGATGAGCCAGCCGAGAGCCCATCTGGAACCCCCAGGCCCCGAGGAACT	1920
QY	1921	gcaagggagcagccagccagagagcccatctggaaccccccaagcccccgaggaacct	1980
Db	1921	GCAGGGAGCAGCCAGCCGAGAGCCCATCTGGAACCCCCAGGCCCCGAGGAACT	1980
QY	1981	gcaagggagcagagccagccgagagcccatctggaaccccccaagcccccgaggaacct	2040
Db	1981	GCAAGGGAGCAGAGCCAGCCGAGAGCCCATCTGGAACCCCCAGGCCCCGAGGAACT	2040
QY	2041	acaagggatgagccagccagagccgagggagagcagcaagatctgaggaacgagagtgagag	2100
Db	2041	ACAAGGATGAGCCAGCCAGCCAGAGCGCGGGGAGAGCAGCAAGATTGCAAGACGAGAGTGAG	2100
QY	2101	tctcttcaccaagtcctggaagaccttaaggaagaagagtgcccgctcgagcttgatctctc	2160
Db	2101	TCTTCTGCCAAGTCTGGGAAGCCTTAAGGAAGAGTGCCCGTGGGGTCTTGTCCTCC	2160
QY	2161	tgctccctgctcagagggctggggccttcggaagctgtctcgagctcccccaagctctgct	2220
Db	2161	TGTCCTCGCTGCAGGGGCTGGGGCTCTCGAGAGCTGCTGCGGCTCTCCCTCAGGCTTGCT	2220
QY	2221	tcgtgacccgtgacccatctgacccagctgctcctctctctctctctgagccatagagcca	2280
Db	2221	TCTGTACCCCTGATACCCATGACCCACATGCTGTGGCTCTCTTGGGGGCACTAATGACAGCCA	2280
QY	2281	ccaagaacccgcgcagagccctcctaagggaagcccaagagcctgcagaaagcctctgacctgctg	2340
Db	2281	CCAGAAGCCCGGAGGCCCTCAGGGAAGCCCAAGGCTGCGAAGAGCCTCTGCGCTGGCTG	2340
QY	2341	tgcttcccccaagctctccctctcgccctgctctgtcttgaattgagcctctgagat	2400
Db	2341	TGCTTCCCCACCCAGCTCTCCCTCGCGCCCTGTCTTGAATTGACCTTCTGAGAT	2400
QY	2401	gggggggccc 2408	
Db	2401	GGGGGGCG 2408	
RESULT 2			
AF172451			
LOCUS	AF172451	2423 bp	mRNA
DEFINITION	Homo sapiens opioid growth factor receptor mRNA, complete cds.	PRI	19-APR-2000
ACCESSION	AF172451		
VERSION	AF172451.1	GI:7595302	
KEYWORDS	.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 2423)		
AUTHORS	Zagzo, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.		
TITLE	Cloning, sequencing, chromosomal location, and function of cDNAs		
	encoding an opioid growth factor receptor (OGFR) in humans		
JOURNAL	Brain Res. 856 (1-2), 75-83 (2000)		
MEDLINE	20143540		
REFERENCE	2 (bases 1 to 2423)		
AUTHORS	Zagzo, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,		
	Pennsylvania State University College of Medicine, 500 University		
	Drive, Hershey, PA 17033, USA		
FEATURES	Location/Qualifiers		
source	1..2423		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="20"		
	/map="20q13.3"		
	/tissue_type="placenta"		
CDS	34..2067		

```
/note="Met-5]-enkephalin binding protein; OGF;  
alternatively spliced"  
/codon_start=1  
/product="opioid growth factor receptor"  
/protein_id="AF64404.1"  
/db_xref="GI:7595303"  
/translation="MDPDCSTWEDEDEDAEDDCEDGEAGARADADAEDEE  
SEEPRAPESSFSQMTGSRNMRATRMCRYPRLVPLVENDCGDPLNLSFRNIR  
FLPNQCFIEDLQMTDNYDLLEDNHSYIOMLFPLRPGVNMHAKPLTLEVEVSS  
QEIOLRVAYEIMLGFYGLIEDRGITVROKROKRONMRSHNLTATRTIK  
SPELSLEHFOALVRFLEETLYRRLPVROSAIDYEMFATRCHOROLYHME  
HFRRCRKFVMPDKLRBKPSPLPHESSRYVEEGSGDPDHSASTGRTGCEH  
SKGGRYDEPQPSRVSPQAPLENSQDEAGHGDEPLSPESKRKLEISRR  
EQPTPEPQPSAEVEXIALNLEGALSQSLRTGTQEVGDPGEAVOPCQPLGAR  
VADVRRRRRVDEAGDAAVASGAGQTLAAGSPAGSHKPKHSENVDEBTERT  
GPKRTGSSPSETGSPSPADDEPASEPSETGPRPAGPAGDEPASEPSETGPR  
AGPAGDEPASEPSETGSPSPAGPPTRDEPASEPSETGPRPAGAGPASEPSETGP  
RPAPAGDEPASEPSETGSPSPAGPPTRDEPARKGEALQDAEVESANSKRP"
```

BASE COUNT 503 a 754 c 843 g 323 t
ORIGIN

Query Match 94.0%; Score 2263.6; DB 11; Length 2423;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 2339; Conservative 0; Mismatches 9; Indels 60; Gaps 1;

```
OY 1 tagattcagggcgccgtgaatttccttaagccagatcgtgacaccccgactgagctccacc 60  
Db 1 TAGAATTCAGCGCGCTGATTTCTTACCGAGCATGAGACACCCCGACTCGCTCCAC 60  
OY 61 tggagagagagagagagatgcgagagacgagagagagactgtcgagagacgagag 120  
Db 61 TGGAGAGAGAGAGAGAGATGCGGAGAGACGCGAGACAGACTGCCAGAGCGCGAG 120  
OY 121 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180  
Db 121 GCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 180  
OY 181 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
Db 181 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
OY 241 agggagatgtgtatgcgacaaactccgactctgtgagagagactgtcagatggg 300  
Db 241 AGGGAGATGTGTATGCGCAACTATCCGATCTGTGGAACGAGACTGTCAATGGG 300  
OY 301 gacacgcacaaactgtgattctcaagaagaatgagatccgcttcgccaacgcgtgttc 360  
Db 301 GACACGCCAAACCTGAGTTCTTACGAATAAGATCCGCTTCCTGCGCCAAACGGCTGTTTC 360  
OY 361 attgagagacattcttcagaactgtgagacaaactatgacctctcttgaggacatctcc 420  
Db 361 ATTGAGAGAACTTCTTCAGAACTGAGCGCAACTATGACTCTCTTGAGAGCAATACATCC 420  
OY 421 tacatcaagtgcgtgttctctctgcgagagacagagagagagatgtgcacagccctc 480  
Db 421 TACATCAAGTGCCTTCTCTCTGCGAGAACAGAGAGTGAATGTGCGAAGCCCTTC 480  
OY 481 acgctcaagagagatgcagaggtgtttaaaagctccagagagatcccaagagcgtgttcgg 540  
Db 481 ACGCTCAGAGAGATGAGAGGTGTTTAAAGCTCCAGAGATCCAGAGACGGCTGTCCGG 540  
OY 541 gctctcagagctatgctggtgcttcaagagatccggtcgtgagagacgagacagggcag 600  
Db 541 GCTCTCAGAGCTATGCTGGGCTTCTTACGGAGATCCGGCTGAGAGACCGAGGAGCG 600  
OY 601 gtgggacgagacagacacagagcgttccagaacccggaactggcgagacacacac 660  
Db 601 GTGGGACGAGACAGACACAGAGCGTTCCAGAACTGGAACCTGGCGAGCCACAC 660  
OY 661 aaccctccatcacacgcatctcaagtcgcgcgtgtgagctgagccttcagagcacttcag 720  
Db 661 AACCTCCCATCACACGCACTCTCAAGTCGCGGTGAGCTGAGACCTCGAGCACTTCAG 720
```

```
OY 721 ggcacactgtccgcttcttctcttgaggagacgcgtgtgcgcgcgcgcgcgcgcgcgcgc 780  
Db 721 GCGCCACTGCTCGGCTTCTTCTGAGAGAGACGCTGTGCGGCGGAGACTCCCGGGGTG 780  
OY 781 cggcaagatgacctgactactcaatgttcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgc 840  
Db 781 CGGCAAGATGACCTTGACTACTCAATGTTCGCGGCTGCGCTCCAGACACCGCGCGCAG 840  
OY 841 ctgtgtcaattcgccttggagacacttccggccgcgcgcgcgcgcgcgcgcgcgcgcgc 900  
Db 841 CTGTGTCAATTCGCTTGAGAGACTTCCGGGCCCGCGCTCCAGAGTTCGCTGCGGCGCCAA 900  
OY 901 gcaagctgtgcgcgcgttcaagcccaagctctgcgcgcgcgcgcgcgcgcgcgcgcgc 960  
Db 901 GCAAGCTGTGCGCGAGTTCAACCCAGCTCTGCGCATCGCTCGAGGCTCCAGAGAG 960  
OY 961 gtgagagagagagagagagagagagagagagagagagagagagagagagagagagag 1020  
Db 961 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
OY 1021 tgtgagccagagacatagcaaggttggggcagaggttggacagagggcccaagcagagc 1080  
Db 1021 TGTGAGCCAGAGCATAGCAAGGTTGGGGCAGAGGTTGACAGAGGGGCCCCAGCAGGAGC 1080  
OY 1081 gtgagagccagagatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140  
Db 1081 GTGAGAGCCAGAGATGCGGAGACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
OY 1141 gggagagatagagagagagagagagagagagagagagagagagagagagagagagag 1200  
Db 1141 GGGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
OY 1201 agccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1260  
Db 1201 AGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 1260  
OY 1261 atgccttgaatttggagaggtgtgtccctcaagcagagagagagagagagagagagag 1320  
Db 1261 ATGCCTTGAATTTGAGAGGTTGCGGCTGCGCTCAGCAGGAGCTCAGAGAGAGAGAG 1320  
OY 1321 gaagtggcggttaagagaccccttgaggagagagagagagagagagagagagagagag 1380  
Db 1321 GAAGTGGCGGTTAAGAGACCCCTTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
OY 1381 aggttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440  
Db 1381 AGGTGCGCGCAGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
OY 1441 ggcgtggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1500  
Db 1441 GCGGTGCGCGAGTGTGTGTGCTCCAGACCTTGCGCTTGCGGCTGCCCTGCCCCATCGGGG 1500  
OY 1501 caccccaagcttgcacacagtgagaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1560  
Db 1501 CACCCCAAGCGTGGACACAGTGAAGCGGGTTGAGAGAGAGACAGAAAGTCAAGCGGGG 1560  
OY 1561 ccccaagagagatgcccttgaggagcccatcggaagcccaagagcccaagcagagact 1620  
Db 1561 CCCAAGAGAGATGCCCTTGAGGAGCCCATCGAGACCCACAGCCCAAGCGGAGAGACT 1620  
OY 1621 gcaagggagagagcagacagaccccatcggaagcccaagagcccaagcgcgcgcgcgc 1680  
Db 1621 GCAAGGGAGAC----- 1629  
OY 1681 acaagagatgagccagcagagagcccatcggaagcccaagagcccaagcgcgcgcgcgc 1740  
Db 1681 -----GAGCCAGCCGAGAGCCCATCGAGACCCAGGCCCCCGGCAAGAGAGACT 1680  
OY 1741 gcaagggagagcagcagagagagagagagagagagagagagagagagagagagagagag 1800  
Db 1681 GCAAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
```

QY	1801	gcaaggagacgaagccagagccaaagatcccatctcgagagaccccaagggcccaagcccgagagact	1860
Db	1741	GCAGGGGACGACCCACGCGCAGACCCATTCGGAGACCCACAGGCCCCACAGCCCCGGCAGAGACT	1800
QY	1861	acaagggatgagccagccagccagacccatctcgagagaccccaagggcccccgcgcgcagagact	1920
Db	1801	ACAAGGGATGAGACCCAGCCAGAGGCCCATTCGGAGACCCACAGGCCCCGGCAGAGACT	1860
QY	1921	gcaagggaacgaagccagcccgagaccccatctcgagagaccccaagggcccccgcgcgcagagact	1980
Db	1861	GCAGGGGACGACCCACGCGCAGACCCATTCGGAGACCCACAGGCCCCGGCAGAGACT	1920
QY	1981	gcaagggaacgaagccagcccgagagaccccatctcgagagaccccaagggcccccgcgcgcagagact	2040
Db	1921	GCAGGGGACGACCCACGCGCAGACCCATTCGGAGACCCACAGGCCCCGGCAGAGACT	1980
QY	2041	acaagggatgagccagcccaagccagggcgaggagagagagatgacagagacgaaggtggag	2100
Db	1981	ACAAGGGATGAGACCCAGCCAGCGCGGGGAGCAGCAGAGTTCTCAGACCGCAGAGTGGAG	2040
QY	2101	tctctgcgaagctctgggaagccttaaggaagagagtgcccgctcgcgctctgtgtcttc	2160
Db	2041	TCTTCTGCCAAGTCTGGGGAAGCCTTAAGGAAGAGAGTGCCTCGGCGCTCTGTGTCTTC	2100
QY	2161	tgtccctctgtgagagagcttgaggcctccggagctgtctcgcgagctctccctcaagcctctgt	2220
Db	2101	TGTCCCTCTGTGAGAGGGGCTGGGGGCTCCGGAGCTCTCGGAGCTCCCTCAGAGCTCTGCT	2160
QY	2221	tcgtaccctgtaccatgagccatgagccagtgctgtgcctctcgttgggccaataagagcca	2280
Db	2161	TGCTACCCCTGTACCCATGACCCACAGTCTGTGCTCTCTGTGGGGCCACTATAGACCCA	2220
QY	2281	ccaagaagccgcgagagccctcagggagaagcccaagggcctgcagaagcctctgagccttgctg	2340
Db	2221	CCAGAAGCCGCGAGGCGCCCTCAGGGAAGCCCAAGGCGCTGCAGAAAGCTCTGGCGTGGCGTG	2280
QY	2341	tgtctcccccaacccagctctccctcgcgccctgtgtctgttaattgaccccttcgaggt	2400
Db	2281	TGTCTTCCCTCACCCACGCTCTCCCTCGCGCCCTGTCTTGTGAATGTGACCCCTTCGAGGT	2340
QY	2401	ggggggcg 2408	
Db	2341	GGGGGGCG 2348	
RESULT	3		
LOCUS	AF109134	2423 bp	mRNA
DEFINITION	Homo sapiens 7-60 mRNA, complete cds.	PRI	10-JAN-1999
ACCESSION	AF109134		
VERSION	AF109134.1	GI:4139227	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 2423)		
FEATURES	1..2423		
FEATURES	source		

CDS	/note="approximately 3kb upstream of COL9A3" 206..2062 /codon_start=1 /product="7-60" /protein_id="AA03737.1" /db_xref="GI:4139228" /translation="MTGSRNMRARMDRCYRHHYPDI.VBERDQNDPPNIS.FFVNEIRF.LPNCFTIEDITQNTWTDLEDNHSTIQMLFLREGGVNMAKPLTIREVEYFKSSQ.EIOLRLVRAVAHAGLLLRDPAGCGRRHGSHGGESTLPALDEPELRSNNILTRILKS.LGEILHFHOAPLVLFRLLESLVRELPGVQSALDYEMFAVG.RHOROLDVFAWH.FRPCKVEMGPQDKLRFKKPSLSLPHLEGSKRYEESBPDPHEASTQRCTGPSHS.KGGRVWGEGPOPFRSEVERODADGLPERSGDGEGEDPEPSPISKEEKRKTELSRE.OPNPBEPQSA.EVEKRIALEGALSOGSLRTGTQGA.GODPEANOPCROLGARV.ADKTKRKRTDEBGADSANAASGAGOTLALAGSABSGHRKACHSENGVEDYERIG.PKESIPESPSETGSPSPAGPDGDEPAISPSETPBPRPAGAGDEPAS.PSETGPRA.GPADDEPAKPSPETGPPSPAGPTDPDEPAESPSETGPPRPGPADDEPAESPSETGPPR.PACGADGEDPASPSETGPPSPAGPTDRPEAPKAEMALDELDAEVSSAKSP"				
BASE COUNT	501 a	759 c	846 g	317 t	
ORIGIN					
Query Match		90.7%	Score 2185;	DB 11;	Length 2423;
Best Local Similarity		96.5%	Pred. No. 0;		
Matches 2300;	Conservative	0;	Mismatches	20;	Indels 63; Gaps 4;
Db	27 gccgcgcatgtgacggccccgcactgctatccacctgggaagaaagcaggagatgcgga	86			
Db	23 GCCGAGCATGACGACGCCCCGACTGCAGCTCCACCCTGGAGAAGACGAGAGATGCCGA	82			
Qy	87 ggaacgcgaagacagagagactgcgaagacgagcgccgcgcgcgaagagaacgcgagacg	146			
Db	83 GGACCGCGGAGACGAGAGACTGGAGAGAGCGGCGCCGCCGCGAGACGCGGAGACG	142			
Qy	147 agggagacgaagacagagagagtgcgagagagccgcgcgcgcgcgcgcgcgcgcgcgc	206			
Db	143 AGGGAGACGAGACGAGAGAGTCGAGAGAACCCCGCGG-GGCGCTGCCCACTGTTCCAGTC	201			
Qy	207 cgaattgcagagttccaaaacttgccagagcccagagagacatgtaagtatctggaca	266			
Db	202 CAGAAATGACAGGGTCCAAGAAATGGCGAGCCACGAGGAGCAATGTAAGTATCGGCAAA	261			
Qy	267 ctatccggaatctgtgtgaaacagaaatgtgaatgagagaccgcgaacatctgaatttcaag	326			
Db	262 CTATCCGGATCTGGTGGAACGAGACTGCAATGGGAGACAAGCCAAACGTGAATTCTACAG	321			
Qy	321 aaatgatgcgcttcctcgtcccaacgctgttcaattgaagacatcttccaagactgac	386			
Db	322 AAATGAGATCCGCTTCCGCCCCAAGGGCTGTTCTCATTTAGGACATTTCTTCAGAACTGAC	381			
Qy	387 ggaacaacatagacctccttggaggaacaacatctccaacatcaagtaggctgttctctgcg	446			
Db	382 AGCACACTATGACCTCTTGAAGACAACTCTCTACTCCAGTAGGCTGTTCTCTCTGCG	441			
Qy	447 agaacacagagatgtaactgtgacatgccaaagccctcacagctcagtcagggagtgtcga	506			
Db	442 AGAACCAAGAGTGAACCTGCATGCCAACCCCTCAACGCTCAGGGAGGAGCGAGGTATA	501			
Qy	507 aagctccagaagatccagagagcggtgtgtccgggctaacg-agctaatgttggtctct	565			
Db	502 AAGCTCCACAGGAGATCCAGGAGCGGCTTGTCCGGGCTTACACACTCATGCTGGGCTTCT	561			
Qy	566 acgggataccggtgtgaaagacacagggacacggtgtggcgcgcgcgcgcgcgcgcgcgcgc	625			
Db	562 ACGGATCCGCTGGAGACCGAGGACCGGGCACGGGTGGCCGAGACAGAACTTACCAGA	621			
Qy	626 agcgcttcacgaacctgtgaacttgcgcagccacaacaacatctcgatatcaagctactca	685			
Db	622 AGCGCTTCACAGAACTGTAAC-T-GCGCACGCCAACAACTCCGATCCACAGCACTCTCA	680			
Qy	686 atgcgcgttgtgagtgtgctctgcagacatctccagggcggcgcgcgtgtcgccttctctg	745			
Db	681 ACTCGCTGGGTAGCTGGGCTTCGAGCACTTCCAGGCGCCGCTGCTGCTTCCTTCGG	740			

```

Qy 746 aggaagacgtggtcgagcgagagctgcgggggtgcgagaggtgacctggaactatca 805
    |||
Db 741 AGGAGAGCCTGTGTGCGCGGAGACTGCCGGGGTGCAGAGTGCCTTGACTACTTCA 800
Qy 806 tgttcgcgtgctgcgagacacgacgacgacgtgtgacttcgtctggagacac 865
    |||
Db 801 TGTTCGCGTGTGCTCCGACACACGCGCGCAGCTGTGTGCACTTCCCTGGGAGACT 860
Qy 866 tccggcgccgctgcaagttcgtctggtgggcccacaagctgcgaggttcaagccca 925
    |||
Db 861 TCGGGCCCCGTGCAAGTTCTGTGGGGGCCCAAGACAACTGCGGAGGTTCAAGCCCA 920
Qy 926 gctctctgcgcataccgtctcgagaggctccagagaagtgtgagagaagaagcccgagg 985
    |||
Db 921 GCTCTCTGCGCATCCGCTCGAGGGCTCCAGAGGTGAGAGGAGAGAGAGAGAGAGAGAG 980
Qy 986 accccgacacgagacgagacacacacacacacacacacacacacacacacacacacac 1045
    |||
Db 981 ACCCCGACACACAGCCAGACACACACACACACACACACACACACACACACACACACAC 1040
Qy 1046 gggagaggtgagacagggggcccgacgagaggtgtgagaccccgagagtgagaccc 1105
    |||
Db 1041 GGGGAGAGGTGACAGAGGGGCCCAAGCCAGAGAGGTGAGAGCCCAAGATGCGGGAGCCC 1100
Qy 1106 tggagagagacgaggggagatgagacaggggccaagggagagatagagccgagaccc 1165
    |||
Db 1101 TGGAGAGAGAGCAGGGGAGATGAGCAGGGGCGCACGGGGAGAGATAGGCCCGAGCCCTTAA 1160
Qy 1166 gcccacaaagagacagagaagagagagagagagagagagagagagagagagagagag 1225
    |||
Db 1161 GCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1220
Qy 1226 ggcagagacacagagatgacacagagtgagagagagagagagagagagagagagag 1285
    |||
Db 1221 AGCCAGAGCCCTCAAGATGCTCTCAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1280
Qy 1286 ccccgacgacgagagagagagagagagagagagagagagagagagagagagagagagag 1345
    |||
Db 1281 CCTTCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
Qy 1346 aggcagatgcagacccctgcgacacacacacacacacacacacacacacacacacacac 1405
    |||
Db 1341 AGGCAGTGCAGACCTCTCGCCCAACCTCTGGAGACAGAGGTGCGCGCAAGGTGAGAGAGAG 1400
Qy 1406 ggaagagaggtgagatgagagagagagagagagagagagagagagagagagagagag 1465
    |||
Db 1401 GGAGAGAGGTGAGTGAAGGTGTGTGGGAGACAGTCTCGGTGGCCAGTGTGTGGCCAGA 1460
Qy 1466 ccttgagcccttgccgggtccctgcgacacacacacacacacacacacacacacacacac 1525
    |||
Db 1461 CTTTGGCCCTTGCCTGGGTCCCTGCCCATCGGAGACACCCCAAGGCTGTGACACAGTGA 1520
Qy 1526 acggaggttgaagagacacagagagagagagagagagagagagagagagagagagagag 1585
    |||
Db 1521 ACGGGTTTGAAGAGGACACAGAGAGGTGAAAGGGGGCCAAAGAGGTACCTCTGGAGAGCC 1580
Qy 1586 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 1645
    |||
Db 1581 CATCGAGAGACCCAGAGCCCAAGCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1640
Qy 1646 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagag 1705
    |||
Db 1625 -----GAGCCGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1640
Qy 1706 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 1765
    |||
Db 1641 CATCGAGAGACCCAGAGCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1700
Qy 1766 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 1825
    |||
Db 1701 CATCGAGAGACCCAGAGCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1760
Qy 1826 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 1885

```

```

Db 1761 CATCGAGAGACCCAGAGCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1820
Qy 1886 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 1945
    |||
Db 1821 CATCGAGAGACCCAGAGCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1880
Qy 1946 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 2005
    |||
Db 1881 CATCGAGAGACCCAGAGCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
Qy 2006 catcgagagaccccgagaccccgacgagagagagagagagagagagagagagagagagagag 2065
    |||
Db 1941 CATCGAGAGACCCAGAGCCCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2000
Qy 2066 gggagagacagagatgcaagagacagagagagagagagagagagagagagagagagagag 2125
    |||
Db 2001 GGGAGAGACAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2060
Qy 2126 aaggaagagatgagccgtcgagagagagagagagagagagagagagagagagagagagag 2185
    |||
Db 2061 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2120
Qy 2186 tccgagagctgctgagagagagagagagagagagagagagagagagagagagagagagag 2245
    |||
Db 2121 TCCGAGAGCTGTGGGGGCTCCCTCAGAGCTGTGCTGTGACCCGATGACCATGACCCAC 2180
Qy 2246 agtgcgtccctcgtgtgagagagagagagagagagagagagagagagagagagagagag 2305
    |||
Db 2181 AGTGTGCTCTCTGTGTGGGACACTATAGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2240
Qy 2306 agcccaagagcctgagagagagagagagagagagagagagagagagagagagagagagag 2365
    |||
Db 2241 AGCCAGAGCCTGCAAGAGCCTCTGCGCTGAGTGTGTCTCCCAAGCTCTCTCCCT 2300
Qy 2366 ggcgcgcgtctgtgtgaatgaccccttcgagagagagagagagagagagagagagagagag 2408
    |||
Db 2301 GCGCCCTGTCTTGTGAATTGACCTTCTGAGTGGGGGGCG 2343

```

```

RESULT 4
AF172452 2363 bp mRNA PRI 19-APR-2000
LOCUS Homo sapiens opioid growth factor receptor mRNA, complete cds.
DEFINITION AF172452
VERSION AF172452.1 GI:7595304
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2363)
AUTHORS Zagon,I.S., Verderrame,M.F., Allen,S.S. and McLaughlin,P.J.
TITLE Cloning, sequencing, chromosomal location, and function of cDNAs
encoding an opioid growth factor receptor (OGFR) in humans
JOURNAL Brain Res. 856 (1-2), 75-83 (2000)
MEDLINE 20143540
REFERENCE 2 (bases 1 to 2363)
AUTHORS Zagon,I.S., Verderrame,M.F., Allen,S.S. and McLaughlin,P.J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA
FEATURES
source location/Qualifiers
1..2363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/issue_type="placenta"
34..2007
/note="[Met-5]-enkephalin binding protein; OGFR;
alternatively spliced"

```

```

/codon_start=1
/product="opioid growth factor receptor"
/protein_id="AA654405.1"
/db_xref="gi:7595305"
/translation="MDPDDCDSTWEDEDEDAEDAEDECDEGEAAGARDADGEDEE
SEEPRAAPSFSGSRMTGRNWRATRMRCYRHHYPDLVRDCDGTDPNLSFYRNEIR
FLPNCGEFIEDLQWMTNNDYLDLEHNYSYIOMLPRLRPGVNMHAKPLTLEVEFKSS
OEIDRLVRAEYELMIGFYGLIEDRGITGVGAQNOYKRONLWMSHNNLRTLRK
SPECISELSEHFAPELVREFLETLVRLRPGVROLLEYFMFAVCRHORROLVHFMF
HERRCKRVWMPDOKLRFRKPSLI.PHLLESGSRKEEGSGDDPHNESTGRCRGPCH
SKGGCRVDEGPRGPRSVGRPDAGPLRGSGRAGGCHGGRPREPLSPKSKRTELISR
EQPTEPRGPQASAEYERIALNLECCALSOSLSLRGITVEYGGDGEVQRCRPLGR
VADYVKKRRKAGDEGASAAVASCQAOTLIALAGSAPSGHPRKSNSEVETEGAT
GPKDTPESPSKETPGSPAGACGDEPESLSETPESGPPADGDEPSESSETPGSP
AGPRLRDEPSPSTPGPRAPGAPDEPAESPSTTPPRAAPGADGDEPAESPSTPGP
SPAPTRNEPKAEAEALQDAEVESSAKSKP"
BASE COUNT      488 a      729 c      825 g      320 t      1 others
ORIGIN

```

[illegible]

Db	721	GGCGACACTGTCCTCCCTTCTCTGAGAGACCTGTGCTGGCGGGAGCTGCGGGGGCTG	780
QY	781	CGGGAAGTGCCTCGGACTCTCATTGATGTTGCGGTGCGTCCGACACCAAGGCGCCAG	840
Db	781	CGGGAAGTGCCTCGGACTCTCATTGATGTTGCGGTGCGTCCGACACCAAGGCGCCAG	840
QY	841	ctggTgcactctgcctctggagacacttcgcgccccgcctgcgaattctgctctggyggcccaa	900
Db	841	CTGGTGCACTTCGCGCTGGAGCACTTCGCGCCCCGCTGCAATTGCTCTGGGGGCCCAA	900
QY	901	gacaagctgcgaggttcaagcccagctctctgcgcatcgcctcgagygctccagyaag	960
Db	901	GACAACTGCGGAGTTCACAGCCCACTCTCTCCSCATCCGCTCGAGGGCTCCAGGAAG	960
QY	961	gtggaagagaaagaaagccccgcgggagccccgcacaaagagcagaaccocaaaggtctgagcc	1020
Db	961	GTGGAAGAGGAAGAAACCCCCGGGGAGCCCCGACCAAGAGCCAGACCAAGGTGTGGACC	1020
QY	1021	TgtagccagagcctagcaagggTggggccaggtgtgacagagggccccagccagcagagc	1080
Db	1021	TGTGGCCAGAGCTATACCAAGGGTGGGGGCGAGGGGTGACAGAGGGCCCCAGCGACGAGC	1080
QY	1081	gtggagcccccaaggaTgcggygacccctctggaagagagccaggggaltgagTcaggggacac	1140
Db	1081	GTGAGAGCCCCAGATGCGGGAGCCCCCTTGAGAGAGCAGAGGGGATGAGGCAAGGGGGCCAC	1140
QY	1141	ggggaaagatagccggygaccccttaagcccccaagagacaaagaaagaaagctgtgagctg	1200
Db	1141	GGGGAAATATAGGCGCGGAGCCCTTAACTCCCAAGAGCAMAAGAAAGAACTGTGAGCTG	1200
QY	1201	agccggggagagccgccccacagagccagagccctcaagatgctcagaggttgaaagag	1260
Db	1201	AGCCGGGGAGAGCGCGGCCCAAGAGCCAGGCGCTCAAGATGCTCAAGAGTGAAGAG	1260
QY	1261	atcgcctgaatttgagagggTgtgcacctcagcccaagggcagctcagagcaggggacccag	1320
Db	1261	ATCGCTGATTTGAGAGGGGTGTGCGCTCAAGCAGGCAAGCTCAGAGCGGGAGCCAG	1320
QY	1321	gaagTgggTcggtcagagaccccttgggagagcaTgTgcagccctgcgcacaacccttggagac	1380
Db	1321	GAAATGGCGCGGTACGAGCCCTATGGGGAGGCAATGACACCTCTCGCGCAACCCCTTGGAGACC	1380
QY	1381	aggTtggccgaacaaagTgaaagaaagcggaaagaaagTbgaTgaaagTaaTcggggaaagTbct	1440
Db	1381	AGGGTGGCCGAAGAGGAGGAAGAGAGAGAGTGAATGAAGGTGCTGAGGACATGTCCT	1440
QY	1441	ggcgTggcccaagTgtgtgtgcacagacctTggcccttgcgggTtccctTgccccactcggg	1500
Db	1441	GGCGTGGCCAGTGTGTGTGCCCAAGACTTGGCCCTTGGCGGGTCCCTGCCCCATCGGGG	1500
QY	1501	caccccaaggtctgnaacagTgaagaaocgggTgtgagagagacacagaaggttcgaaocggg	1560
Db	1501	CACCCCAAGGCTGACACACTGTGAAGACGGGGGTGAGAGAGACACAGAAAGTGTGAACGGGG	1560
QY	1561	cccaagaagagTaccctctggagcccatctggagaccccaagccccagccccagccccagagact	1620
Db	1561	CCCAAGAAAGGTACCTCTGGGAGCCCATCGAGAACCCCAAGGCCCCAGGCCA-----	1611
QY	1621	gcagggagcagccagccaaagaccatctgagagccccagccccagccccagccagagact	1680
Db	1612	-----	1611
QY	1681	acaaaggaTgagccagccgagagccatctgagagccccagcccccccgccagagact	1740
Db	1612	-----GCAAGGACT	1620
QY	1741	gcagggagcagccagccagccagagcccatctggagaccccaagcccccccgcccgagagact	1800
Db	1621	GCAAGGACCAAGCCAGCCGAGACCCATCGAGAACCCCAAGGCCCCCGCCGCGCAGAGACT	1680
QY	1801	gcagggagcagccagccaaagatcccatctgagagccccagccccagccccagccagagact	1860

Dh 1681 GCAGGGAGAGCCAGCCAGAGCCCATGAGACCCAGGCCAGCCGAGACCT 1740
Qy 1861 acaagagatgagccagccgagagcccatcgagagcccccagcccgccgagagact 1920
Db 1741 ACAAGGATGAGCCAGCCAGAGCCCATGAGACCCAGGCCAGCCGAGACCT 1800
Qy 1921 gcaagggagagccagccagagcccatcgagagcccccagcccgccgagagact 1980
Db 1801 GCAGGGAGAGCCAGCCAGAGCCCATGAGACCCAGGCCAGCCGAGACCT 1860
Qy 1981 gcaagggagagccagccagagcccatcgagagcccccagcccgccgagagact 2040
Db 1861 GCAGGGAGAGCCAGCCAGAGCCCATGAGACCCAGGCCAGCCGAGACCT 1920
Qy 2041 acaagagatgagccagccagagcccgagagccatgagagcccccagcccgagagact 2100
Db 1921 ACAAGGATGAGCCAGCCAGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Qy 2101 tcttcgcaagctctgagagcccttaaggaagagtgccgctcgagctctgctcttc 2160
Db 1981 TCTTCTGCCAAGTCTGGGAGACCTTAAGGAAGAGTGCCTCGGCTTGTGCTCTCC 2040
Qy 2161 tcttcgcaagctctgagagcccgagagccctcgagagctctgagagctctgct 2220
Db 2041 TGTCTCTGCTCAAGGGCTGGGCTCGGAGCTGCTGCGGCTCCCTCAGGCTCTGCT 2100
Qy 2221 tctgtaaccctgtagcccaagcccaagctgctgctctgctgagagccatagagccca 2280
Db 2101 TCGTGAACCCGCTGAGCCATGAGCCAGACTGCTGCTGCTGAGGAGGAGGAGGAGGAG 2160
Qy 2281 ccaggaagccgagagcccgagagcccgagagcccgagagcccgagagcccgagagcc 2340
Db 2161 CCAGGAAGCCCGAGAGCCCGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Qy 2341 tgcttcgccagccagctctccctcgagagcccgctgtgttaagagccctcgagag 2400
Db 2221 TGCTTCCCAACCCAGCTCTCCCTCGGCTCTGCTTGTAAATGAGCCCTTCTGAGT 2280
Qy 2401 gggggggc 2408
Db 2281 GGGGGGCG 2288

RESULT 5
HS885L7
LOCUS HS885L7 160241 bp DNA PRI 09-JUL-2000
DEFINITION Human DNA sequence from clone RP5-885L7 on chromosome
ACCESSION AL035669
VERSION AL035669.43 GI:8979786
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 160241)
Smith, M.
Direct Submission
Submitted (09-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clone.requests@sanger.ac.uk
On Jul 8, 2000 this sequence version replaced g1:8919619.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'

Feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
EMBL: Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HEP/Chr20>
RP5-885L7 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-885L7 The true left end of clone RP4-563E14 is at 140192 in this sequence. The true right end of clone RP11-93B14 is at 44512 in this sequence.
Location/Qualifiers
1. 160241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.2-13.33"
/clone="RP5-885L7"
/clone_lib="RPCI-5"
2. 64
/note="LIMA9 repeat: matches 6210. 6254 of consensus"
65. 353
/note="Alu10 repeat: matches 1. 287 of consensus"
354. 397
/note="LIMA9 repeat: matches 6254. 6302 of consensus"
4198. 5157
/note="12 copies 80 mer 61% conserved"
complement(4223..5153)
/note="random repeat: region contains two forced joins and single clone. Assembly consistent with restriction digest."
4229. 5114
/note="443 copies 2 mer cc 59% conserved"
4231. 5166
/note="234 copies 4 mer tccc 60% conserved".
4240. 5151
/note="19 copies 48 mer 61% conserved"
4248. 5159
/note="38 copies 24 mer 61% conserved"
4264. 5163
/note="15 copies 60 mer 61% conserved"
4286. 5152
/note="17 copies 51 mer 59% conserved"
4308. 5171
/note="27 copies 32 mer 61% conserved"
4317. 5156
/note="21 copies 40 mer 62% conserved"
4344. 5142
/note="17 copies 47 mer 59% conserved"
4348. 5145
/note="42 copies 19 mer 60% conserved"
4431. 5144
/note="21 copies 34 mer 60% conserved"
4456. 5148
/note="21 copies 33 mer 60% conserved"
4486. 5151
/note="9 copies 74 mer 69% conserved"
4530. 5087
/note="18 copies 31 mer 71% conserved"
4533. 4822
/note="5 copies 58 mer 77% conserved"
4720. 5137
/note="11 copies 38 mer 61% conserved"
4833. 5064
/note="4 copies 58 mer 79% conserved"
5051. 5164
/note="19 copies 6 mer tcttc 63% conserved"

FEATURES

source

[illegible]

Db	6075	GACCCAGAGCCCCCGCCGCGAGACCTTGACAGGGGACGACGCCAGCCAGACCCATCGGA	6134
QY	1953	gaccccaagcccccccgcccgagagacctgcagaggagcaagccagagccatcgga	2012
Db	6135	GACCCAGAGCCCCCGCCGCGAGACCTTGACAGGGGACGACGCCAGCCATCGGA	6194
QY	2013	gaccccaagcccccccgcccgagagacctacaaggatgagccaagcgaggagagc	2072
Db	6195	GACCCAGAGCCCCCGCCGCGAGACCTTGACAGGGGACGACGCCAGCGGGGAGGC	6254
QY	2073	agcaagattgcagagcgcgaaggtgtgaattcttctgcgaagtcttggaaagccttaagaaa	2132
Db	6255	AGCAAGATTGCAGAGCGAGAGGTGAAGTCTTGCCAAAGTCTGGAGAGCCTTAAGGAAA	6314
QY	2133	gagagcgccctgtgaggtcttgtctctctgtccctgtcgcagaggagctggagcctccgag	2192
Db	6315	GGAGTGCCTGTGAGGTGGCTTTGGTCTCTCTGTTCCTCTGCTGCGAGGGGCTTGGGCTCCGGAG	6374
QY	2193	ctgctgcgggctcccccataagagctctgtcttcgtgacccgtgacccatgacccagtgctg	2252
Db	6375	CTGCTGCGGGCTCCCTCAGGCTGTGCTTCGAGACCCGTGACCATGACCAAGTGTG	6434
QY	2253	gcctctcttggggcactatagcagcaccagaagcgcgagagccctcgaagagcccaa	2312
Db	6435	GCCCTCTTGGGGCCACATATAGCACCCACCAAGAGCCCGAGGCCCTCAGGAGAGCCCAA	6494
QY	2313	ggcctgcagaagccctcctgcgctgtgtgtcttcccccacacagctctccctcgcgcc	2372
Db	6495	GGCCTGCGAAGACCTCCTGCGCTGGCTGTGCTTCTTCCCAACCAAGCTTCCCTTGGCCCC	6554
QY	2373	tgtctttaaattgaccttcttgcagtgaggagggc	2408
Db	6555	TGCTTTGTAATTGACCTTCTGAGTGGGGGGG	6590
RESULT	7		
AF172450			
LOCUS	AF172450	1676 bp	mRNA
DEFINITION	Homo sapiens	opioid growth factor receptor mRNA, complete cds.	19-APR-2000
ACCESSION	AF172450		
VERSION	AF172450.1	GI:7595300	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 1676)		
	Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.		
	Cloning, sequencing, chromosomal location, and function of cDNAs encoding an opioid growth factor receptor (OGFR) in humans		
	Brain Res. 856 (1-2), 75-83 (2000)		
JOURNAL	20143540		
MEDLINE	2 (bases 1 to 1676)		
REFERENCE	Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.		
AUTHORS	Direct Submission		
TITLE	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,		
JOURNAL	Pennsylvania State University College of Medicine, 500 University		
	Drive, Hershey, PA 17033, USA		
	Location/Qualifiers		
FEATURES			
source	1..1676		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="70"		
	/map="20q13.3"		
	/tissue_type="placenta"		
	34..1419		
	/note="[Met-5]-enkephalin binding protein; OGFR; alternatively spliced"		
	/codon_start=1		
	/product="opioid growth factor receptor"		
	/protein_id="AA64403.1"		
	/db_xref="GI:7595301"		
	/translation="MDPPDCDSWEEDEEDAEADAEDEGAEDEGAAGRADADAEDEE		
CDS			

SEEPARBPSSFSQSMTCRNRATRDCKRYRNPDLVERDCNGDTPLNLSFYRNEIR
 FLNPGCFIEDLQWMTDNDYDLLEDNHSYIOMLEPLRBPVNMHAKPLTLREVEVSS
 QDEIDRLVATLEMLGFIYEDIEDRGITGVRQNTKRONLWNRSHNRLTTRILK
 SPELSTLEHFAPLVREFLETLVRELPVROGALDYFMEAVROROLVHPAME
 HFRPKCFVWGPQDRLFRKPSLPHLEGRKVEEYSGSDDHENSTOGRCEBH
 SKGGRVDEGQPSRVEPDQPLERSGOAGHGDREDEPSPRSKSRKRIELSR
 EOPTREPOSASEVENKIALNBECALSSQSLRGTIOEVGGQDPGEASCPCCRWGLR
 SCCPLSPSSAS"

BASE COUNT 340 a 501 c 575 g 260 t
 ORIGIN
 Query Match 56.0%; Score 1349.4; DB 11; Length 1676;
 Best Local Similarity 98.5%; Pred. No. 1.4e-189;
 Matches 1362; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 tagaattacgagcgccgtgnaattcttaagccgagcatgagacagcccgactgcagctccac 60
 Db 1 TAGAATTACGCGCGCGCGGAATTCATGCCGAGATGACGCCGCCGACTGCGACTCCACC 60
 QY 61 tggagagagagagagagagatgagagagagagagagagagagagagagagagagagagag 120
 Db 61 TGGAGAGAGAGAGAGAGAGATGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 121 ggc 180
 Db 121 GCCCGCGCGCGCGAGGAGCGCGGAGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 181 ggc 240
 Db 181 GCG 240
 QY 241 agggagacatgtgtagatctgcgcacaactatccgagtcgtgtgagacagagctgagatg 300
 Db 241 AGGACATGTGTAGATCTGCGCAACTATCCGATCTGTGAGAGAGAGAGAGAGAGAGAG 300
 QY 301 gacacgc 360
 Db 301 GACACGCGCAACTGAGTTCTACAGAAATGAGATCCGCTTCTGCGCAACGGCTGTTC 360
 QY 361 attgagacatctcttaagaactgagagacacatgactctcttgaagacacatctctc 420
 Db 361 ATTGAGACATCTTCTGAGACTGAGACACTATGACTCTCTTGAAGACACTATCTCC 420
 QY 421 taatcctgagctgttctcctctgcgcagagagagagagagagagagagagagagagagag 480
 Db 421 TACATCAGAGGCTGTCTCTGCGAGAACAGAGAGTGAATGGCATGCCAAGCCCTC 480
 QY 481 acgctcaaggagagctgagctgtttaaagctccagagagagagagagagagagagagag 540
 Db 481 ACGCTCAGGAGAGTCAAGTGTTTAAAGCTCCAGAGATCCAGAGACGGCTGTTCGG 540
 QY 541 gctcaagagctatgctgagcttcttaagagagagagagagagagagagagagagagagag 600
 Db 541 GCTTACAGACTATGCTGGGCTTCTACGGGATCCGGCTGAGAGACCGAGCGGCGCG 600
 QY 601 gtcgagc 660
 Db 601 GTGGGCGGAGCAAGAACTACCAAGAGCGCTTCCAGAACTGAACTGGCGAGCGACCAAC 660
 QY 661 aacctcgc 720
 Db 661 AACCTCCGATACACAGCATCTCAAGTCGCGGTGTAGCTGAGACCTCGAGCACTTCCAG 720
 QY 721 ggc 780
 Db 721 GCGGCACTGTGCTTCTCTGAGAGAGAGCGCTGTGCGCGGAGCTGCGGCGGTG 780
 QY 781 cgc 840
 Db 781 CGGCGAGAGTGCCTGTGAGACTTATCTATGCTGCGCTGCGAGACACAGCGCGCGAG 840
 QY 841 ctgtgtgcacatcgcttgagagacacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900

Db 841 CTGTGTGCACATTTGCTTGGAGGACATTCGCGGCCCTGCTCAAGTTCGCTGGGGGCCCCAA 900
 QY 901 gacaagctgc 960
 Db 901 GACAAGCTCGGAGGTTCAAGCCAGCTCTGCGCGCATCCGCTCGAGGCTCCAGAGAG 960
 QY 961 gtcgag 1020
 Db 961 GTGAG 1020
 QY 1021 tctgagcagagacatagcag 1080
 Db 1021 TCTGAGCAGAGACATAGCAG 1080
 QY 1081 gtcgag 1140
 Db 1081 GTGAG 1140
 QY 1141 gggag 1200
 Db 1141 GGGAG 1200
 QY 1201 agc 1260
 Db 1201 AGCGCGCGGAG 1260
 QY 1261 atcgcctgaaattgag 1320
 Db 1261 ATCGCTGAAATTGAG 1320
 QY 1321 gaagtgagc 1380
 Db 1321 GAAGTGCGCGGATGAG 1380
 QY 1381 agg 1383
 Db 1381 CGG 1383

RESULT 8
 AF172449
 LOCUS
 DEFINITION
 Homo sapiens clone 127 opioid growth factor receptor mRNA, complete cds
 ACCESSION
 AF172449
 VERSION
 AF172449.1 GI:7595298
 KEYWORDS
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 1232)
 Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
 Cloning, sequencing, chromosomal location, and function of
 encoding an opioid growth factor receptor (OGFR) in humans
 Brain Res. 856 (1-2), 75-83 (2000)
 JOURNAL
 MEDLINE
 20143540
 REFERENCE
 2 (bases 1 to 1232)
 Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
 Direct Submission
 Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
 Pennsylvania State University College of Medicine, 500 University
 Drive, Hershey, PA 17033, USA
 FEATURES
 source
 1..1232
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="20q13.3"
 /clone="127"
 /tissue_type="placenta"
 34..1212

CDS

gene	complement(1316. .2065) /gene="H75" complement(1316. .2065) /gene="H75" /note="similar to HHV6 U75, HCMV UL103, HVS gene 42, EBV BBRF2, EHV1 55, VZV 53, HSV1 UL7" /codon_start=1 /product="H75" /protein_id="AAC40325.1" /db_xref="GI:2769713" /translation="MKSLLKLELETSVDVNTLHVTKILAVIKIDKCVSLARSLVNI TVGDGIMFLEHEDGTMLNGLEYKTKICEKELGFOGIIIIIDSEDTLOELRLNFORK RLILHMVDPPEEMLGGLVLENLPLKOSTLKLRLVLPVLSPTKILFNTDQ KLVTCLRHIFENEASELHKVPMIRLYNMKNMTHLEVELEYNTKRSNFIILSLE SROLODSSLOVIOIATOPAOIFYSKNEDTS" complement(11974. .3962) /gene="H76" complement(11974. .3962) /gene="H76" /function="possible viron protein" /note="similar to HHV6A U76, HCMV UL104, HVS gene 43, EBV BBRF1, EHV1 56, VZV 54, HSV1 UL6" /codon_start=1 /product="H76" /protein_id="AAC40326.1" /db_xref="GI:2769714" /translation="MHRASNLISVSGSMRMNOSSGRRPRSLSDNATLSTINS ILGAEMLSKNLSYLPNNBEIDMTIPSEDMITFIELHGHKNEFGQITHNLRDS AVLKQIAYGAQAOLNSVSIQIHDWMKRVNRPNNKRLSPQDYFSVWAHAKQV ILGDSINIINFILQSIDNSHNRVYDIICVGPENRTPTAPDLYNLQOVSXKLI HDIVRHONITVTPILLGLSSVILPDEFNIRKIFRNSEQISCFENKKAIAFYTSY VIRRRMLTTPPLAHSPELKHNSLRHOKKCOLNFPPIKVLTLATDYNKIMM IKERKNSDAKSLIKPELLNSDKSKIGIIPSEGTIOETPSIIDQNTLMNRGP RKSSAIDTGERDVADELKQIITKCEMOIQIDMETITLTKTNOMETKIKDLSLE TNNDCDRINPDLDELNLISRLNLVQRLPFTSVSIDTRVANSFQIYIPDTQY ADKRIDQLEMEYMERFLRNKNVNOGSEITYSNLSIEIIVFPLRLNLTNLS IPEEFLLSGEILAIYESKIKHYLRLVYRELNOISIEVYNLTQHPENNEPIFDS NIESPNEENLEKIKRIRNLRIOHLRPNYPGDDQ" 3757. .6231 /gene="H77" 3757. .6231 /gene="H77" /function="helicase, helicase-primase complex" /note="similar to HHV6A U77, HCMV UL105, HVS gene 44, EBV BBLF4, EHV1 57, VZV 55, HSV1 UL5" /codon_start=1 /product="H77" /protein_id="AAC40327.1" /db_xref="GI:2769715" /translation="MSTISLFGGRYDNKFLIMSSAPKTELIVDKVASLSERLEGRL PEWFRIMPETEFENSEPADALCIGIDERAOPPLPFRKALVYTRAGAGKTRISOT LAANLDCIVATITISIAONLSVVLNRSKSAQVKITFTFGNSHVSMSEROSTIAD ERSIOIQOKODLSYWNVISDIARLGAVACKRELPDCE\$SVIYIDAGVILRH LHVVEFYWYNALYKTPYLENGIIVPCIVGSPQSNALVTSNPLOKNDVYRGID VLSALICDVLKSCYCEVDNMNIIIPVNNRCADHAFGDFLKHIEGLPKRELLIYVO FVKASYIRNDPMNEIETRLFLISNEYKPNRSIHEQVETNRRNLTFFVPIYIKRK TFEDYKSEIGNSLEIEIPWPKSNHRLANTISQPADOLSKTVQEEVLEDGSEETL ITCLHAKRNSISGVTSIKASIVGSGTYEKVELQSDIFIKTSCDPTIAYSTL SGLWFGMYSCFCESTTPVLMKIKIMKPSIEFLSEMSRMSRDQVOTETBERYDF GLVDDGSLMDLLEIDPCDPFEFRYKSLPLNLSFEISLITTKFDLFI\$FALI OMHTKKGFKTLVTVYNNVNSRKQCEIYSHLKSFPGLTYAIPANNYILEGTNN VHLGTDKOLPOLIYKKGGLPRIVYKDMGFPSTVDNNSKFPVDVNGOSHICCTVY ATYSKVSMTITKSGSLIQKVALIDFGSDPKMLKLSIYVGM\$RVTDPNNLTMMVPLR LNTENDNFIAHVYKALKNENTMLIF" 6495. .6722 /gene="CB5RH" 6495. .6722 /gene="CB5RH" /codon_start=1 /product="CB5RH" /protein_id="AAC40342.1" /db_xref="GI:2769730" /translation="M\$SAKTNESLPRSTRMYSGGNLKNAGQFSLLSRELVNSYPDVNS	gene	GGGHDGYL\$INRSSFIETTEVILGIGVRIR" 8179. .8406 /gene="CB6RH" 8179. .8406 /gene="CB6RH" /codon_start=1 /product="CB6RH" /protein_id="AAC40343.1" /db_xref="GI:2769731" /translation="MLHNSKCIH\$ICIRVCILKLVHICSSRFSIRCFAYEYSLIPN TSDGRQAFIAYFGKVALVADAIRKGFRRG" 8829. .9866 /gene="H79" 8829. .9866 /gene="H79" /function="possible replication protein" /note="similar to HHV6A U79, HCMV UL112, spliced in vitro replication protein" /codon_start=1 /product="H79" /protein_id="AAC40328.1" /db_xref="GI:2769716" /translation="MYAEERGYSFDNVIOAYEOLII\$OSLHLNRPFEDNCGFTEFLAD SGCIEFSGKIM\$IMYWTSETDSGLTVDIGDEKRCARYRARGLLCSIT\$ISO NTG\$ERILFVSHENGLOITTEVYIAKVSSEPLRNLDLKEPEKEKECRAIDRKKH DDEHRRSGKOKERKRYVEDIDKREDEKIKOEKRRNDEKRDPKKDEPDGKEVAY VNFPIIG\$KSM\$FPLNSICFVYISEPPEKRO\$H\$ETKRNLEQ\$HEDGIAPT\$ITE VNA\$VAGALS\$PCV\$IDNHEDQ\$HDELDRVYAOVGVLS\$KPRPSLESLCV\$KADLF LIDDEPRRSIQ" 9838. .10449 /gene="H80" 9838. .10449 /gene="H80" /function="possible replication protein" /note="start-codon ggg; similar to HHV6A U80, HCMV UL113, spliced in vitro replication protein" /codon_start=1 /transl_except="(pos:9838. .9840,aa:met) /product="H80" /protein_id="AAC40329.2" /db_xref="GI:7687902" /translation="MTNREGLSMEDNYGNFOLKSLDEOLRRLRVASSGEYVERSL\$KL KE\$LDVYKDLIN\$VLECADVY\$KSL\$TK\$HIEOKIOIV\$DCY\$V\$VCIK\$EPI DM\$EVEFETTONAR\$VRO\$RRTV\$GSDG\$IGQ\$RYSIGONR\$RARGR\$RRRN SNLNLTQNSAIVIDD\$SETENFENAG\$NEDLATTILETL" complement(10493. .11260) /gene="H81" complement(10493. .11260) /gene="H81" /note="similar to HHV6A U81, HCMV UL114, HVS gene 46, EBV BBRF3, EHV1 61, VZV 59, HSV1 UL2" /codon_start=1 /product="uracil-DNA glycosylase" /protein_id="AAC40330.1" /db_xref="GI:2769718" /translation="MALLQMLD\$HND\$EKNYENL\$TIDDOH\$LF\$G\$INPML\$T\$Q\$K LEITHLK\$VYKLDND\$R\$H\$TV\$PSSDN\$HANS\$FLCK\$P\$DVK\$VYI\$IGHDPY\$P\$R\$CG L\$AGTY\$EKC\$SIP\$SLN\$IFELERS\$IPN\$SP\$NG\$LN\$S\$RG\$VILLNS\$IT\$F\$V\$G\$L P\$M\$E\$F\$G\$W\$T\$L\$Y\$K\$IS\$R\$S\$EDMS\$IV\$FL\$M\$G\$K\$H\$AR\$K\$SYL\$ID\$A\$K\$H\$Y\$LE\$S\$A\$H\$P\$S K\$Y\$A\$A\$R\$P\$F\$G\$CH\$F\$V\$T\$N\$FL\$F\$E\$H\$G\$K\$D\$P\$IN\$M\$IL\$N\$E" complement(11336. .12088) /gene="H82" complement(11336. .12088) /gene="H82" /function="gH-gL complex, infection, cell fusion" /note="gH: similar to HHV6A U82, HCMV UL115, HVS gene 47, EBV BBRF2, EHV1 62, VZV 60, HSV1 UL1" /codon_start=1 /product="gHycoprotein L" /protein_id="AAC40331.1" /db_xref="GI:2769719" /translation="MELLF\$VMS\$ILL\$LF\$SK\$AMP\$LP\$D\$HNS\$F\$Y\$E\$K\$LD\$D\$C\$IA\$V\$IN\$CTR SEVPLLEPIYQ\$APVY\$NEDV\$M\$ILL\$N\$P\$P\$TK\$K\$F\$S\$IM\$T\$N\$E\$T\$LD\$F\$ILL\$Q\$N\$PE\$QLR
------	---	------	--

gene /protein_id="BAA78215.1"
/db_xref="GI:4995982"
/translation="MSSVFEDEPSTHSPTSRSPSTRSPSTRSPSTRSPSTRSP
CGRSPSPVARVRIASVAALTHRSLEIRIIVSRALANPSGDEPTRMSPMPDPCG
VGTSTAAVVGGSVLPGADYSRTVWTCACREGAATRGSGIAGPLAPRAGSRVCR
PSSACVALS"
5025. .5336
/gene="DR6"
5025. .5336
/gene="DR6"
/note="84.5% identical to DR6 gene of strain U1102 of
HIV-6"
/codon_start=-1
/protein_id="BAA78216.1"
/db_xref="GI:4995983"
/translation="MTTRHTQDGRGRIAIRDGAATLARAARPEWLLARGSPSL
VGYTSRHRERHILPMPRYWCLLEHPDYRDARSATVWGHWGMPPIHYRVSQDCC
P"
6512. .7150
/gene="DR7"
6512. .7150
/gene="DR7"
/function="transforming activity"
/note="42.2% identical to DR7 gene of strain U1102 of
HIV-6"
US22 gene family"
/codon_start=-1
/product="transactivator"
/protein_id="BAA78217.1"
/db_xref="GI:4995984"
/translation="MRAARCLASNTVEFARGLQDPVAEEDLELPRRMGTVN
RLFVPIAADAEDHLLTVGGICQTHAVSCLELTVDPMTAVPGVMTLTVARQY
RLMPARAPRLHLVAGHPDGPIVCAVLSBERETQPSPIGSGVTLGVPPGPR
EYETAMVITYWDRBCYRSGPIARSAVWRTSPCGMAAGHENTLDVFGG"
complement(7236. .7706)
/gene="DRH2"
complement(7236. .7706)
/gene="DRH2"
/note="ORF unique to HST strain"
/codon_start=-1
/protein_id="BAA78218.1"
/db_xref="GI:4995985"
/translation="MAPPTPGVYIKGECGALSSSPDLCDSPCVGIGRCYSMSARG
RCLIDRKVFPFLCGSNTERKHYCCVSPILFLGLAGRDVTATVTHPTHR
RHIDISTAVLFLRLRTPHFFPVLLFELIWMFLASCMPIAISRLRGV"
7928. .8662
/gene="DR8"
7928. .8662
/gene="DR8"
/note="13.6% identical to DR8 gene of strain U1102 of
HIV-6"
/codon_start=-1
/protein_id="BAA78219.1"
/db_xref="GI:4995986"
/translation="MQQTVTRDDKGRLLANOLGKKRKORCKILTKETSLLPPTPST
TLPPLRNHFCGRFSRVRVQOTRBRHATTSMAVAGGAGSIRGADYKSMNGIARV
PSSRLAIIAROGKPLAOSITLITLITLITLITLITLITLITLITLITLITLIT
LTLITLITLITLITLITLITLITLITLITLITLITLITLITLITLITLITLITL
CAPRGSGPLPPLPKNGRIGTGGWRV"
complement(8292. .8807)
/gene="LJ1"
complement(8292. .8807)
/gene="LJ1"
/note="16.6% identical to L1 gene of strain U1102 of
HIV-6"
/codon_start=-1
/protein_id="BAA78220.1"
/db_xref="GI:4995987"
/translation="MAIAAVSLFRLILVCSAALPASISLRAPSPFRSHPARGC
CTFARLAIHPIHTLARSIGGGGSGIPAGKGTGDTKGLGVRGRGFTFHAPLAVD
VNTAPLEAARRALVDGLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLG
LGLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLGVLG
89329. .9384

Query Match	Best Local Similarity	Score	DB	Length
Matches 273; Conservative	54.68%	136.8	161573	0; Gaps 0; Indels 0;
QY 1564	aaagaagtlacccctggagcccatcggagagcccaagcccaagccgaagaccca	1623		
DB 129493	ACAGAGCCTCCAGCAGACGACCTCCAGACAGGCTCCAGAGAGCCTCCAGAGCCTCC	129434		
QY 1624	gggagagccagccagccagcccatcggagagcccaagcccaagcccgagagaccta	1683		
DB 129433	GCAGAGCCTCCAGCAGACGACCTCCAGACAGGCTCCAGAGGCGCTCCAGAGCCTTCA	129374		
QY 1684	aggagatgagccagccagagcccatcggagagcccaagcccccagcccgagagaccta	1743		
DB 129373	GCAAGGCTCCAGCAGACGACCTCCAGAGAGCCTCCAGAGAGCCTCCAGAGGCGCTTCA	129314		
QY 1744	gggagagccagccagccagcccatcggagagcccaagcccccagcccgagagaccta	1803		
DB 129313	GCAGAGCCTCCAGCAGAGGCGCTCCAGAGGCGCTCCAGAGAGCCTCCAGAGGCGCTCG	129254		
QY 1804	gggagagccagccagccagagatcccatcggagagcccaagcccccagcccgagagaccta	1863		
DB 129253	GTAAGGCTTCCAGCAAACTTCCAGCGGAGGCTTCAAGCAAAAGCTTCCAGCAGGCGCTTCA	129194		
QY 1864	aggagatgagccagccagagcccatcggagagcccaagcccccagcccgagagaccta	1923		
DB 129193	GAAAGAGCCTCCAGCAGAGGCGCTCCAGAGAGCCTCCAGAGAGCCTCCAGAGAGCCTCA	129134		
QY 1924	gggagagccagccagccagagcccatcggagagcccaagcccccagcccgagagaccta	1983		
DB 129133	GCAGAGCCTCCAGCAGAGGCGCTCCAGAGAGCCTCCAGAGGCGCTTCCAGCAGAGCCTTCA	129074		
QY 1984	gggagagccagccagccagagcccatcggagagcccaagcccccagcccgagagaccta	2043		
DB 129073	GCAAGAGCCTCCAGCAGAGCCTCCAGAGGCGCTCCAGAGGCGCTTCCAGAGAGCCTTCA	129014		
QY 2044	aggagatgagccagccagagcc	2063		
DB 129013	GCAGAGCCTCCAGCAGAGCCTCCAGAGCCTCCAGAGCCTCCAGAGCCTCCAGAGCCTTCA	128994		

[illegible][illegible]

gene
6413. .9409
/gene="HDFR2"
CDS
6413. .8887
/gene="HDFR2"
/note="putative helicase; similar to HCMV UL105,
Swiss-Prot Accession Number P16736"
/codon_start=1
/protein_id="AA68468.1"
/db_xref="GI:662100"
/translation="MSISLFGGRYDNKFLNMSAPKIELIYDKVASLSERLERGL
PEDWFRHIMDEPTEFSEFADALCIGIDEAFOALPLPEFALLVYTAGAGKNSLOT
LAALDCIATATSIANON.SVLNRSKQVYTIKTRFNSHMSRSTYLAND
ERSTIOOKODISYWNVYSDIADRALGAVCKTEKPELCESSVLIIDAGVTLNHI
LHYVFFWYFNALYKTPLEDGIVPCIVCGSPQSNALVYSFNLQNKDVKGID
VLSALICDVLVSKYCEVDNMWILFVNKRCADAFGDFLKHIEGFLPELLEYVDQ
FVKPASYIRNPMIEETRLFLSHNEVKNFRLSHQVEYTNRNLLFVPPVFLINK
TFEDYKSEIGNFSEIEPWFKSNHRLNTYSDPADDLSTVOLEIIVLEDGSVEETL
ITCHKIIRNSISGVTSKIKASVTFGSPQEKERKEVLOSLDIFTEKSCQOTIHASFL
SGLMFGMYSCCSTTPPEVMEIKNIKPSIEFLESEMSRSPVQVTERIDRDF
GIYNDGLSDVLEIDPCGDPFTFRYSKPLNLSLSEFELSLYTFKDIISRAIL
OKHTKGFETLVTYNNRNVSRKQGEIYSHLSKPYGMILYAIIPANNYTLGTYND
VHLGTDKQLPQLIKRGLPLVINDMGIISVLDNNVEKFDIVVGQSFHLCTYDY
ATVSKVSMITTKSQGLSIQKVAIDFGSDPRNLKLSIYVMSRVTDPNNLINNVPLR
LNYENDNFIAPIHVKALKMENTWLF"

polya_signal

gene
/evidence="not-experimental"
complement(9325. .10351)
/gene="EDLF5"
9404. .9409
/gene="HDFR2"
/evidence="not-experimental"
complement(10022. .10351)
/gene="EDLF5"
/codon_start=1
/product="unknown"
/protein_id="AA68469.1"
/db_xref="GI:662101"
/translation="NMMLAYFLYCRQLLAVALVLEPPRLCGNTIRFYSYHLPLSTCH
ISDSFYLTSRTSNLLFCHLYCHOKSHOKSHOKKTGTGLKLSWTINIKMENSSP
VTLSTTA"
11477. .12652
/gene="EDRF1"
11477. .12511
/gene="EDRF1"

gene
CDS
11477. .12511
/gene="EDRF1"

/note="putative phosphoprotein; similar to HB6 CB7R,
GenBank Accession Number U14772 and HCMV UL112, Swiss-Prot
Accession Number P16768"
/codon_start=1
/protein_id="AA68470.1"
/db_xref="GI:662102"
/translation="MYAERAVGSFDDVMEVYQVIESIOLKRLHFGSGCLIEFLD
SGTCTFCGGWISMTCTSDTNSMGLVLDIGDKTKITNANGFLCSKTSISIQ
NTEGRDRLITYSQENKQLQITFTLVTKFEDHDIRNGDPKICIEFEKCRALDRKH
DDEHRRSGKQEKREKVEDTKDDKDDKQEKRRNDKEDKOPDKESDGKYCIYV
VNFYFSCVTVDFLNSICFVAISELPKEROXYHDMKRNLEQSHEDGITLSTFLY
NGAVGALPFCISIDNHEDQOHELDKRAVAQVGSPPRSGSLCYSKADLEI
LGDEPRRSIQ"
12483. .13180
/gene="EDRF2"
12483. .13079
/gene="EDRF2"
/note="putative phosphoprotein; similar to HCMV UL113,
Swiss-Prot Accession Number P17151"
/codon_start=1
/function="replication"
/protein_id="AA68471.1"
/db_xref="GI:662103"
/translation="GYNREGLSNEDNYGNFLNLSLEQLRAKLVASSGDIYERSLKL
KECLDNVKNLIRNEKADYVTPSKLCTKHIEPKQIVFSDCVRPVQCEIKPFIDV

Query Match

4.2%; score 101.8; DB 72; length 21191;

Best Local Similarity 51.2%; Pred. No. 1.3e-06;
Matches 238; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
QY 1579 ggaagcccatcggaagaccaccagccaccagcaggaactcgagggagagccagcc 1638
DB 19496 ggagctccagcagagagcctccagcagagagcctccagcagagagagctccagc 19437
QY 1639 aagaccatcggaagaccaccagccaccagcccgaggaactgaagggatgagccagcc 1698
DB 19436 gagctccagcagagagcctccagcagagagcctccagcagagagcctccagc 19377
QY 1699 ggaagcccatcggaagaccaccagccaccagcccgaggaactcgagggagagccagcc 1758
DB 19376 ggagattctagcagagagcctccagcagagagcctccagcagagagagcctccagc 19317
QY 1759 ggaagcccatcggaagaccaccagccaccagcccgaggaactcgagggagagccagcc 1818
DB 19316 gagctccagcagagagcctccagcagagagcctccagcagagagagcctccagc 19257
QY 1819 aagatcccatcggaagaccaccagccaccagcccgaggaactgaagggatgagccagcc 1878
DB 19256 gagctccagcagagagcctccagcagagagcctccagcagagagagcctccagc 19197
QY 1879 ggaagcccatcggaagaccaccagccaccagcccgaggaactcgagggagagccagcc 1938
DB 19196 ggagattctagcagagagcctccagcagagagcctccagcagagagagcctccagc 19137
QY 1939 ggaagcccatcggaagaccaccagccaccagcccgaggaactcgagggagagccagcc 1998
DB 19136 aagctccagcagagagcctccagcagagagcctccagcagagagagcctccagc 19077
QY 1999 ggaagcccatcggaagaccaccagccaccagcccgaggaactcgagggagagccagc 2043
DB 19076 aagctccagcagagagcctccagcagagagcctccagcagagagagcctccagc 19032

Search completed: October 12, 2000, 17:56:51
Job time: 34787 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 08:17:04 ; Search time 17897.3 seconds
(without alignments)
556.841 Million cell updates/sec

Title: US-09-431-843B-4
Perfect score: 2290
Sequence: 1 tagaattcagcgccgctga.....ccttcgagtggtggggcgcg 2290

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 218378903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_cm:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_pl3:*
34: gb_pl4:*
35: gb_pl5:*
36: gb_htg1:*
37: gb_htg2:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: gb_htg8:*

44: gb_htg9:*
45: gb_htg10:*
46: gb_htg11:*
47: gb_htg12:*
48: gb_htg13:*
49: gb_htg14:*
50: gb_htg15:*
51: gb_htg16:*
52: gb_htg17:*
53: em_ba1:*
54: em_ba2:*
55: em_htg1:*
56: em_htg2:*
57: em_htg3:*
58: em_htg4:*
59: em_htg5:*
60: em_htg6:*
61: em_htg7:*
62: em_hum3:*
63: em_hum4:*
64: em_hum5:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v11:*
73: gb_v12:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pl6:*
92: gb_pl7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2276.4	99.4	2363	11 AF172452	AF172452 Homo sapi
2	2200.4	96.1	2423	11 AF172451	AF172451 Homo sapi
3	2137.2	93.3	2483	11 AF172453	AF172453 Homo sapi
4	2118.6	92.5	2423	11 AF109134	AF109134 Homo sapi
5	1555.6	67.9	160241	35 HS88517	AL035569 Human DNA
6	1542.8	67.4	6670	11 AF112980	AF112980 Homo sapi
7	1351	59.0	1676	11 AF172450	AF172450 Homo sapi
8	977.4	42.7	1232	11 AF172449	AF172449 Homo sapi
9	698.4	30.5	2208	12 AF156878	AF156878 Rattus no
10	106.4	4.6	30752	72 HHU92288	U92288 Human herpe
11	103.2	4.5	161573	72 AB021506	AB021506 Human her
12	101.4	4.4	162114	72 AF157706	AF157706 Human her


```

Qy 841 ctgtgtgacattcgcgtggaagcacttccggcccgctgcaagtctgtctgtggggcccaa 900
    |||
Db 841 CTGGTGCACTTGGCTGGGAGCACTTCGGGCCCGCTGCAAGTTCCTGGGGGCCCCAA 900
Qy 901 gacaagctgaggaggttaagccagactctctgtcccaatccgctcgagggtctccaggag 960
    |||
Db 901 GACAAGCTGCGGAGGTTAAAGCCACAGCTCTTGGCGCATCCGCTCGAGGGCTCAGAGAG 960
Qy 961 gtgtgaggaggaagaaagcccgaggagcccgaaacagagggcagaccaggtctgagacc 1020
    |||
Db 961 GTGAGGAGGAGAAAGGCCCGGGGACCCGACCCAGAGGCGACACCCAGGCTGGAGCC 1020
Qy 1021 tgtgtgccaagagcaatagaaggtgtggggagaggttgagcagaggggccccagccagagc 1080
    |||
Db 1021 TGTAAGCCAGAGCATACCAAGGCTGGGGGCAAGGTGAGCAGAGGGGCCCCAGCCAGGAGC 1080
Qy 1081 gtgtgagccccaagagatgagagagcccttgagagagagagccagagggagtgagagccac 1140
    |||
Db 1081 GTGAGGCCCAAGAGTGGGGGACCCCTGAGAGAGAGCCAGGGGGATGAGGCGGGGGCAC 1140
Qy 1141 ggggagaatagggccggagcccttaagccccaagagagagagagagagagagagagagctg 1200
    |||
Db 1141 GGGGAATATAGGCCGGGAGCCCTTAAGCCCAAGAGAGCAAGAGAGAGAGAGAGCTGGAAGCTG 1200
Qy 1201 agccgagggagagagagcccgccacagagagagagagagagagagagagagagagagag 1260
    |||
Db 1201 AGCCGCGGGAGAGAGAGCCGCCACAGGGGCAAGGCTCAGAGTCTCAGAGTGTGAGAGAG 1260
Qy 1261 atgagcttgaattgtggaggggtgtgcctcagccagagagagagagagagagagagagag 1320
    |||
Db 1261 ATGAGCTTGAAATTTGGAGGGGTGTGCCCTCAGCCAGAGGCGAGCTCAGAGAGAGAGAGAG 1320
Qy 1321 gaagtgaggcgagcagagagccctggggagagagagagagagagagagagagagagagagag 1380
    |||
Db 1321 GAAGTGAGGCGGTAGAGAGCCCTGGGGAGAGAGAGTGCAGAGCTCAGCCCAAGCTTGGAGAGC 1380
Qy 1381 aaggtgtgagcgaagaaggttgagagagagagagagagagagagagagagagagagagag 1440
    |||
Db 1381 AGGAGTGGCGACAAAGGTGAGAGAGCGGAGAGAGAGTGAAGAGGTACTGGGGAGACTGCT 1440
Qy 1441 ggggtggcaggtgtgtgtggccagagacttgagagagagagagagagagagagagagagag 1500
    |||
Db 1441 GCGGTGGCAGTGTGTGTGTGCCAGAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1500
Qy 1501 caaccacaaagcttgagacaggttgagagagagagagagagagagagagagagagagagagag 1560
    |||
Db 1501 CAACCCCAAGCTGAGACAGTGAAGAGGGGTTGAGAGAGAGACAGAGAGTGAAGAGGGG 1560
Qy 1561 cccaaagaaaggttaacccttgaggagccatcgagagagagagagagagagagagagagagag 1611
    |||
Db 1561 CCCAAAGAAAGGTACCCCTGGAGCCCATCGAGAGAGCCCAAGGCCCAAGAGAGAGACT 1620
Qy 1612 ----- 1611
Db 1612 GCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1680
Qy 1612 ----- 1620
Db 1612 ACAAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1740
Qy 1621 ggaagggagagagagagagagagagagagagagagagagagagagagagagagagagagag 1680
    |||
Db 1621 GGAAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1740
Qy 1741 gcaagggagagagagagagagagagagagagagagagagagagagagagagagagagagag 1800
    |||
Db 1741 GCAAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1800
Qy 1681 gcaagggagagagagagagagagagagagagagagagagagagagagagagagagagagag 1740
    |||
Db 1681 GCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1800
Qy 1801 gcaagggagagagagagagagagagagagagagagagagagagagagagagagagagagag 1860
    |||
Db 1801 GCAAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1860
Qy 1741 acaagggatgagagagagagagagagagagagagagagagagagagagagagagagagag 1800
    |||
Db 1741 ACAAGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1860
Qy 1861 ACAAGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1920
    |||
Db 1861 ACAAGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1920
Qy 1801 gcaagggagagagagagagagagagagagagagagagagagagagagagagagagagagag 1860
    |||
Db 1801 GCAAGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1860

```

```

Db 1921 GCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 1980
Qy 1861 gcaagggagagagagagagagagagagagagagagagagagagagagagagagagagagag 1920
    |||
Db 1861 GCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2040
Qy 1921 acaagggatgagagagagagagagagagagagagagagagagagagagagagagagagag 1980
    |||
Db 1921 ACAAGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2040
Qy 1981 tcttctgcaagctcgtggagagagagagagagagagagagagagagagagagagagagag 2040
    |||
Db 1981 TCTTCTGCAAGTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2100
Qy 2041 tgtccctgtgagagagagagagagagagagagagagagagagagagagagagagagagag 2100
    |||
Db 2041 TGTCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2160
Qy 2101 ttctgtacccgttgagagagagagagagagagagagagagagagagagagagagagagagag 2160
    |||
Db 2101 TTCTGTACCCGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2219
Qy 2220 TTCGTGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2279
    |||
Db 2220 TTCGTGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2279
Qy 2161 accagagagagagagagagagagagagagagagagagagagagagagagagagagagag 2220
    |||
Db 2161 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2239
Qy 2221 gtgtcttcccaagagagagagagagagagagagagagagagagagagagagagagagagag 2280
    |||
Db 2221 GTGTCTTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACT 2299
Qy 2281 tggggggggcgg 2290
    |||
Db 2281 TGGGGGGGCGG 2409

```

```

RESULT 4
AF109134
LOCUS Homo sapiens 2423 bp mRNA complete cds. PRI 10-JAN-1999
DEFINITION AF109134
VERSION AF109134.1 GI:4139227
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2423)
AUTHORS Liu,J., Brewton,R.G., Takanosu,M., Wood,B.M. and Wayne,R.
TITLE Cloning of 7-60: A human gene from chromosome 20q13.3 which encodes
a novel intracellular protein motif repeated seven times
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2423)
AUTHORS Liu,J., Brewton,R.G., Takanosu,M., Wood,B.M. and Wayne,R.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1998) Cell Biology, University of Alabama at
Birmingham, 1670 University Blvd. VHS02, Room 605, Birmingham, AL
35294, USA

```

FEATURES

```

source .1..2423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/feature="approximately 3kb upstream of COL9A3"
/feature="2062"
/codon_start=1
/product="7-60"
/protein_id="AAD03737.1"
/db_xref="GI:4139228"
/translation="MTGSRNRRATRCMKRYRHNYPDLVERDNGDTFNLSTFYNEIRF
LPNGCFIDIIQMTNDLLEHNSYIOWLPRLREGVNMHAPLRLREVEYKSSQ
EIQRILVRAIYAHAGLRLDPAGGPRHGHGPGSTLPELPALEPRLSHNNILRITRLKS

```

CDS

```

1..2423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/feature="approximately 3kb upstream of COL9A3"
/feature="2062"
/codon_start=1
/product="7-60"
/protein_id="AAD03737.1"
/db_xref="GI:4139228"
/translation="MTGSRNRRATRCMKRYRHNYPDLVERDNGDTFNLSTFYNEIRF
LPNGCFIDIIQMTNDLLEHNSYIOWLPRLREGVNMHAPLRLREVEYKSSQ
EIQRILVRAIYAHAGLRLDPAGGPRHGHGPGSTLPELPALEPRLSHNNILRITRLKS

```

LGELGLHFQAPLVRFLEESLVRELPGVROSALDYFMFVAVGCRHOROLVHFMWH
FRPRCKFVMGPODLRFRKPSLPHLESGSKVEEESPDPDEHLESTOGCTEPRES
KGGGRVDEGPDRSVEPODAPLERISOGDEAGGEMRPELSRSEKSRKRLSTRE
QPRPERPOSSEVEKITALINEGALSOGSRTORTOYEGSDPREANOPCQPLGARY
ADKTRKRKVDGADSNVAGSAGQTLALAGSPBPSHPFAGSSENGVEDTIGRTG
PKERTPSSETPGPSPAGPAGDPAESPSTPPRAGPAGDPAESPSTPPRPA
GPADDEPAKPTSETPGPSPAGPDRDEPAESPSTPPRPADEPAESPSTPPR
PAPGAGEPAPSPSETPGPSPAGPDRDEPAESPSTPPRPADEPAESPSTPPR
BASE COUNT 501 a 759 c 846 g 317 t
ORIGIN *

Query Match 92.5%; Score 2118.6; DB 11; Length 2423;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 2241; Conservative 3; Mismatches 17; Indels 64; Gaps 5;

QY 27 gccagacatgagacaccccgactgcgaactccacctggagagagagagagagatgcyga 86
DB 23 GCCGAGCATGAGACGACCCGACTGCGACTCCACTGGGAGGAGAGAGAGATGCGGA 82

QY 87 ggaagcgagc 146
DB 83 GGACCGGAGC 142

QY 147 agggagc 206
DB 143 AGGGAGACAGATC 201

QY 207 cagaatgacagc 266
DB 202 CAGAAATGACAG 261

QY 267 ctatccgagatctggtgagagagagagagagagagagagagagagagagagagagc 326
DB 262 CTATCCGATCTGgtgagc 321

QY 327 aaatgagatccgctctccgccaagcgtgtttcatltgagagacattcttcagagagc 386
DB 322 AAATGAGATCCGCTTCCGCCCAAGGCTGTTCATTAGAGACATCTTTCAGAACTGAGAC 381

QY 387 ggaacaatagacactccttgagagagacactcctacatcagtggtcttctcctgagc 446
DB 382 AGACAACATATACCTCTTGAGAGACATCTCTATCATCAGTGGCTGTTCCTCTCGC 441

QY 447 agaagc 506
DB 442 AGAACCAAGTAA 501

QY 507 aagctccagagagatcagagagagagagagagagagagagagagagagagagagc 565
DB 502 AAGCTCCAGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCT 561

QY 566 acggagatccgagc 625
DB 562 ACGGATCCGAG 621

QY 626 agcgcttcacaaactgagagagagagagagagagagagagagagagagagagagc 685
DB 622 AGCGCTTCACAAACCTGAACT-GCGACAGCCACAAACCTCCGATCCACAGCACTCCA 680

QY 686 agtcgagctgtgagc 745
DB 681 AGTGCTGTGGGTGAGGTGGCTCCAGACATTCAGAGCCCGCTGGTTCGGCTTTCCTGG 740

QY 746 aggaagagctgtgagagagagagagagagagagagagagagagagagagagc 805
DB 741 AGGAGAGAGCTGTGGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 800

QY 806 ttttcgagctgtgagagagagagagagagagagagagagagagagagagagc 865
DB 801 TGTTCGCGCTGTGGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 860

QY 866 tcgagcccgctgcaagctgtctgtgagagagagagagagagagagagagagc 925

DB 861 TCCGGCCCCGCTGCAAGTGTCTGTGGGGGGCCCCAAGCAAGACTCGGAGGTCAAGCCA 920

QY 926 gctctctgcccacatcgctcgaagaggtctccagaaagtgtgagagagagagagagc 985
DB 921 GCTCTCTCCCATCCGCTCCGAGGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980

QY 986 acccgagc 1045
DB 981 ACCCGAGACAGGTG 1040

QY 1046 gggcagaggtgagc 1105
DB 1041 GGGCAGAGGTGAG 1100

QY 1106 tgaagc 1165
DB 1101 TGAG 1160

QY 1166 gccccaagc 1225
DB 1161 GCCCAAG 1220

QY 1226 agccagagccctcagagagagagagagagagagagagagagagagagagagc 1285
DB 1221 AGCCAGAGCCCTCAG 1280

QY 1286 cctcagcagc 1345
DB 1281 CCTCAGCAG 1340

QY 1346 aggcaggtgagc 1405
DB 1341 AGGAG 1400

QY 1406 ggaagc 1465
DB 1401 GGAG 1460

QY 1466 ccttgagcccttgagagagagagagagagagagagagagagagagagagagc 1525
DB 1461 CCTTGAGCCCTTGAG 1520

QY 1526 acggaggttggagagagagagagagagagagagagagagagagagagagc 1585
DB 1521 ACGGGGTGAG 1580

QY 1586 catcgagc 1606
DB 1581 CATCGAG 1600

QY 1607 -----gccagcagagagagagagagagagagagagagagagagagagc 1645
DB 1641 CATCGAG 1700

QY 1646 catcgagc 1705
DB 1701 CATCGAG 1760

QY 1706 catcgagc 1765
DB 1761 CATCGAG 1820

QY 1766 catcgagc 1825
DB 1821 CATCGAG 1880

QY 1826 catcgagc 1885
DB 1881 CATCGAG 1940

QY 1886 catcgagc 1945

Db	1941	CATCGGAGACCCAGGGCCCCAGGCCGCGCAGAGACTACAAGAAGATTGACCAGCAAGCGG	2000
QY	1946	gggaaggcagcaagatgttgcaakcgcagaagtgtgactcttcgccaaagcttgggaagccct	2005
Db	2001	GGGAAGGCACGAGATGTTCAGAACGACGAAAGGTGGAGTCTTTTGCCAAATCTGGGAAGCTTT	2060
QY	2006	aaggaagaagtaccgctcgacgctctgtgctcctcctgtccctgtgcaagggtctggagcc	2065
Db	2061	AAGGAAGAGAGTCCCGTCGGGCTCTTGTGTCCTCTGCTCCCTGCTGCAAGGGGCTGGGGCC	2120
QY	2066	tccggagcttgtctgctggggtcccctaagctctgtctgtgacccgtyaacccaatgacca	2125
Db	2121	TCCGGAGAC-TGCTGCGGGCTCCCTCAGGCTCTGTCTCGAGACCCGCGACCATGACCCA	2179
QY	2126	caattgtgacctctctgtggggcaccatagacaacaccaaagaacggcgagggcccttaagg	2185
Db	2180	CAGCTGTGGCTCTCTGTGGGGCACACTAATGACACCAACCAAGCCCGAGGCCCCTAAGG	2239
QY	2186	aagcccaaggtcctcagaagcctcctgtgctgtgtgtttcctccacaacgaagctcccc	2245
Db	2240	AAGCCCAAAGGCTCGAGAACCTCTTGCCCTGCTGTGCTGCTCCACCAAGCTCTCCCC	2299
QY	2246	tgcggccctgtcttttaattgacctctctgtgagtgaggggcgcg	2290
Db	2300	TCGGCCCCCTGTGTATAATGACCTTCTGTGAGTGGGGGGCGG	2344
RESULT	5		
LOCUS	HS885L7	160241 bp	DNA
DEFINITION	Human DNA sequence from clone RP5-885L7 on chromosome	PRI	09-JUL-2000
ACCESSION	AL035669		
VERSION	AL035669.43		GI:8979786
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 160241)		
AUTHORS	Smith,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone Requests: clonerequests@sanger.ac.uk		
COMMENT	On Jul 8, 2000 this sequence version replaced gi:8919619. During sequence assembly data is compared from overlapping clones. Where differences are found they are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EMBL; SW:, SWISSPROT; Tr:, TRMBL; Wp:, WORMEP; Information on http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence was generated from part of bacterial clone confits of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/RGP/Chr20 RP5-885L7 is from the library RPI-5 constructed at the Roswell Park Cancer Institute by the laboratory of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: PCYPAC2 This sequence is the entire insert of clone RP5-885L7 The true left end of clone RP4-563E14 is at 140192 in this sequence. The true		

```
FEATURES                                right end of Clone RP11-93B14 is at 44512 in this sequence.
```

SOURCE	Location/Qualifiers
1..160241	/organism="Homo sapiens"
/db_xref="taxon:9606"	
/chromosome="20"	
/map="q13.2-13.33"	
/clone="RP5-885L7"	
/clone_1ib="RPC1-5"	
2..64	
/note="LMMA9 repeat: matches 6210..6254 of consensus"	
65..353	
/note="AluYo repeat: matches 1..287 of consensus"	
354..397	
/note="LMMA9 repeat: matches 6254..6302 of consensus"	
4198..5157	
/note="12 copies 80 mer 61% conserved"	
complement(4223..5153)	
/note="Random repeat. region contains two forced joins and single clone. Assembly consistent with restriction digest."	
4229..5114	
/note="443 copies 2 mer cc 59% conserved"	
4231..5166	
/note="234 copies 4 mer tccc 60% conserved"	
4240..5151	
/note="19 copies 48 mer 61% conserved"	
4248..5159	
/note="38 copies 24 mer 61% conserved"	
4264..5163	
/note="15 copies 60 mer 61% conserved"	
4286..5152	
/note="17 copies 51 mer 59% conserved"	
4308..5171	
/note="27 copies 32 mer 61% conserved"	
4317..5156	
/note="21 copies 40 mer 62% conserved"	
4344..5142	
/note="17 copies 47 mer 59% conserved"	
4348..5145	
/note="42 copies 19 mer 60% conserved"	
4431..5144	
/note="21 copies 34 mer 60% conserved"	
4456..5148	
/note="21 copies 33 mer 60% conserved"	
4486..5151	
/note="9 copies 74 mer 69% conserved"	
4530..5087	
/note="18 copies 31 mer 71% conserved"	
4533..4822	
/note="5 copies 58 mer 77% conserved"	
4720..5137	
/note="11 copies 38 mer 61% conserved"	
4833..5064	
/note="4 copies 58 mer 79% conserved"	
5051..5164	
/note="19 copies 6 mer ttcttc 63% conserved"	
5934..6864	
/note="LIM4 repeat: matches 4509..5488 of consensus"	
6974..7049	
/note="4 copies 19 mer 80% conserved"	
7224..7451	
/note="6 copies 38 mer 98% conserved"	
7892..8074	
/note="MER6A repeat: matches 68..210 of consensus"	
8674..8966	
/note="AluY repeat: matches 5..296 of consensus"	
10876..10962	
/note="MIR repeat: matches 86..162 of consensus"	
11974..12025	
/note="13 copies 4 mer ctca 73% conserved"	
14394..14478	
/note="5 copies 17 mer 77% conserved"	

[illegible][illegible]

Dp	5335	GAGCCAGGGGGATGAGGCAGGGGGCCACGGGGAGATGAGGCCGGAGCCCTTAAGCCCA	5414
Qy	1173	agagaagaagaagaagaagctctgagcttgacccgcygagacagccagccacagaagccag	1232
Dp	5415	AGAGAGCAAGAAAGAGAGAGTGGAGGTGAGCGCGCGGAGCAGACCGCCACAGACCCAG	5474
Qy	1233	ccctcagaatgtcctcagaagtgtagaagaatcgctgaatttgagaggggtgtgtccctcag	1292
Dp	5475	CCCTCAAGATGCGCTCAGAGAGGTGAGAAAGATCCGCTCGAATTGGAGGGGTGTGCCCTCAG	5534
Qy	1293	ccaagagagagctcaagagccgagagcccaagaatgagggcggtcaagagccctggagagagcgt	1352
Dp	5535	CCAGAGGAGAGCTCAGAGACGGGGAGCCAGAGAGTGGCGGTCAAGACCTGGGGAGGCGGT	5594
Qy	1353	gcagccctctgcgcgaacccctctggagacccaagtgtygcgcgaacaaggttgagaaacgagaa	1412
Dp	5595	GCACCCCTCGCCGCCAACCCCTGGGAGCCAGAGGTGGCGGCAGAAAGGTAGAGAAAGCGAGAA	5654
Qy	1413	ggtgatgataggtgtctgtggagacagtcgtcgtgtgtgcagtggtgtgtgtgtgtgtgtgtgtgt	1472
Dp	5655	GGTGATGAGAGGTGTCTGGGAGAGTGTGTGGTGGCGGACGATGGAGGTGCCAGAGCTTGTGC	5714
Qy	1473	cccttgccgggtctccctctgtccctcctcgtgggcacccgaagcttgacacagttagaacygt	1532
Dp	5715	CTTGGCGGGGTCCCTTCCTCCCATCGGGGCGACCCCAAGCGCTGACACAGTAGAAGAGGGGT	5774
Qy	1533	tgaagagagacacagaagaagtgtgaacccggagcccaagaagaagttaccccttggaagccatcga	1592
Dp	5775	TGAAGAGAGACACAAAGATGAAAGGGGGCCCAAGAAAGGTACCCCTGGAGCCATCGGA	5834
Qy	1593	gaacccagagcccca-----1606	1606
Dp	5835	GACCCCAAGGCCCCAGGCCACAGACTCGTAGGGAGCAGAGCGGGCCAGAGGCCATCGGA	5894
Qy	1607	-----gccagcaggaacctgtcaagtgagcagccagccgagagcccatcga	1652
Dp	5895	GACCCCAAGGCCCCCGCCCAAGAGACTCGTAGGGAGCAGAGCGGAGAGGCCATCGGA	5954
Qy	1653	gaacccagagcccccgcccgcccgacaggaacctgtgaagggagcagagccgagagcccatcga	1712
Dp	5955	GACCCCAAGGCCCCCGCCCGGACAGACTCGTAGGGAGCAGAGCCCAAGACCCATCGGA	6014
Qy	1713	gaacccagagccccagcccccgccagagactctcaagggatgaagcagccagagcccatcga	1772
Dp	6015	GACCCCAAGGCCCCCGCCCGGACAGACTCTAGAGGATGAGCCAGCCGAGAGCCCATCGGA	6074
Qy	1773	gaacccagagcccccgcccgcccgacaggaacctgtcaagggagcagcagccgagagcccatcga	1832
Dp	6075	GACCCCAAGGCCCCCGCCCGGACAGACTGTGAGGGAGCAGAGCCAGAGGCCATCGGA	6134
Qy	1833	gaacccagagcccccgcccgcccgagagactctcaagggagcagccagccgagagcccatcga	1892
Dp	6135	GACCCCAAGGCCCCCGCCCGGACAGACTGTGAGGGAGCAGAGCCAGAGGCCATCGGA	6194
Qy	1893	gaacccagagccccagcccccgccagagactctcaagggatgaagcagccagagccgagagagc	1952
Dp	6195	GACCCCAAGGCCCCCGCCCGGACAGACTTACAGAGGATGAGCCAGGCCAAGCGGGAGAGCG	6254
Qy	1953	agcagagattgcagkacgacagaagtgtagtctctgcacaaagtcttggaagcccttaagaa	2012
Dp	6255	AGCAGAGATTGAGAGACCAAGTAGAGTGGAGTCTTGCCAAAGTGTGGGAAGCCTTAAGGAA	6314
Qy	2013	ggagatgccgttcgagcgtctgtgtctctctgtctctgtctctgtctcagggagcgtcgag	2072
Dp	6315	GGAATGCCCGTTCGCGCTTGTCCTGTCTGTCTCTCTCAAGGGGCTGGGGCTCGCGAG	6374
Qy	2073	ctgtctctcgagcctccctcagaagcttgagctctgtgtcgtgagccgttgagccatgaaccaagctct	2132
Dp	6375	C-TGCTGCGGCTCTCTCAAGGCTTGCTTGTGAGACCCGTGACCCATGACCCACATGACCTAGTCT	6433
Qy	2133	ggactctctgtgtggccactataagcscacacagaagccgagagccctcctaaggaagcca	2192

QY	DB	6434	GGCCTCTGTGGGGCCATTATAGACAGCCACCAGAAAGCCGACAGGCGCTCTAGAGAAAGCCA	6493
QY	2193	aggacctgagaagacctctctgagcctgctgctgctctccaccaccagctctcccttgccgc	2252	
DB	6494	AGGCTTGAGAAAGCCTCTCTGGCTGCGTGTGTCTTCCACACCAAGCTCTCCCTGGCGCC	6553	
QY	2253	ctgtcttggtaaatgacccctctgagtgaggggcg	2290	
DB	6554	CTGTCTTTGTAATATGACCTTCTGAGTGGGGGCGG	6591	
RESULT	7			
AF172450	AF172450	1676 bp	mRNA	PRI
LOCUS				19-APR-2000
DEFINITION				Homo sapiens oiploid growth factor receptor mRNA, complete cds.
ACCESSION				AF172450
VERSION				AF172450.1
KEYWORDS				GI:7595300
SOURCE				human.
ORGANISM				Homo sapiens
REFERENCE				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS				Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE				1 (bases 1 to 1676)
JOURNAL				Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
MEDLINE				Cloning, sequencing, chromosomal location, and function of CDNAS
REFERENCE				encoding an oipoid growth factor receptor (OGFR) in humans
AUTHORS				Brain Res. 856 (1-2), 75-83 (2000)
JOURNAL				20143540
MEDLINE				2 (bases 1 to 1676)
REFERENCE				Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
AUTHORS				Direct Submission
JOURNAL				Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
FEATURES				Pennsylvania State University College of Medicine, 500 University
SOURCE				Drive, Hershey, PA 17033, USA
				Location/Qualifiers
				1..1676
				/organism="Homo sapiens"
				/db_xref="taxon:9606"
				/chromosome="20"
				/map="20q13.3"
				/tissue_type="placenta"
				34..1419
				/note="([Met-5]-enkephalin binding protein; OGFR;
				alternatively spliced"
				/codon_start=1
				/product="oipoid growth factor receptor"
				/protein_id="AA64403.1"
				/db_xref="GI:7595301"
				/translation="MDPPDCSTWEDEDEDAEDADEDCDEGEAGAGARDADDEDEE
				SEEPRAPSFSFOSRMTGSRNMRATRMCRIRHNYPLTVRDCDGDGPNIASFYRNER
				FLPGRCFIEDILQWMTNPNYLLDENHSYIOMLPRLRPYNNMAKPTLREVFYFS
				OEIDRIVRAITELMGLGTRLEDRGGTIGRAONTOYKRQNLNMRSHNLRITRIIK
				SPCLSLSEHQAPDLKRLFRFPSSLPHPLDEGSRKVEEGSGDPDHEASTGRTGCFH
				HFPRCLSEVQPGDRLKRLFRFPSSLPHPLDEGSRKVEEGSGDPDHEASTGRTGCFH
				SKGGRTGPGDPSVPEODAGPLERSGDGEAGHGDRREPLSPKSKRKRLLSRR
				EQPTREPOSASEVEKIALNLESCALSQSLSIRGTQTEVGGDPEAGSESCPCRWGLR
				SCCCLPBGSGAS"
BASE COUNT	340 a	501 c	575 g	260 t
ORIGIN				
Query Match	59.0%	Score 1351;	DB 11;	Length 1676;
Best Local Similarity	98.6%	Pred. No. 4.7e-189;		
Matches 163;	Conservative	0;	Mismatches 20;	Indels 0;
			Gaps	0;
QY	1	tagaatcagcgccgcggaattctcagccgagcatgacgagaccgccgactgcgactcac	60	
DB	1	TGAAATTCAGCGCGCGCGTGAATTCTTAGCGGAGCATGACGACACCCGACTGGAGTCCAC	60	
QY	61	tggagagagagcagagagatgctgagagagacgctgagagagagagactgcgagagcgag	120	
DB	61	TGGAGAGAGAGACGAGAGAGATGCTCGGAGAGACCGGAGAGACGAGACTGCGAGAGCGCG	120	

[illegible]

Db	1201	AGCGCGGGAGAGCCGCCACAGAGCCAGGCCCTC	CAAGTGCCTCAAGSTGGAGAAG	1260
QY	1261	atcgctctgaattcgaagggtgtgcctcaagccag	gcagcctcgaagcgcgaagcgcag	1320
Db	1261	ATCGCTCTGAATTGAGGGGTGTGCCCTC	CACGACAGGGGAGGCTCAGAGGGGAGCCAG	1320
QY	1321	gaatctggcgcgtcgaagaccctctggagagcagt	gcagccctgcgcgcaaaccccttggagac	1380
Db	1321	GAACTGGCGGCTCAGGACCTGGGAGGCGTCTCT	CCCTCGTGCAGAGGAGCTGGGGCTTC	1380
QY	1381	agg	1383	
Db	1381	CGG	1383	

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1232)	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P. J.	Cloning, sequencing, chromosomal location, and function of encoding an opioid growth factor receptor (OGFR) in humans	Brain Res. 856 (1-2), 75-83 (2000)

AUTHORS
Zacon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE
Direct Submission
JOURNAL
Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA

FEATURES	SOURCE
Location/Qualifiers	1..1232
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="20"	
/map="20q13.3"	
/clone="127"	
/tissue_type="placenta"	
34..1212	
/note="[Met-5]-enkephalin binding protein; OGRF;	

BASE COUNT	256	a	369	c	403	g	204	t
ORIGIN								

Query Match	42.7%	Score 977.4	DB 11	Length 1232
Best Local Similarity	99.4%	Pred. No. 2.8e-134		
Matches 981	Conservative 0	Mismatches 6	Indels 0	Gaps 0

QY 1 tagaattcagcgccgctgaattcttagccgagcatgacgacccgcgactcgactcaacc 60

Db 1 TAGAATTCAGCGCGCGCTGATATCTAGCCGAGCATGACGACACCCCGACTGCGACTCCACC 60
QY 61 tgggaagagagcgagagagatgagagagagagagagagagagagagagagagagagagag 120
Db 61 TGGAGAGGAGCAGAGAGATGCGGAGAGACCGGAGCAGAGAGACTGCGAGACCGCGAG 120
QY 121 gcccgc 180
Db 121 GCCCGCGCGCGGAGGAGCGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
QY 181 ggc 240
Db 181 GCGGCGCGCGCGGAGGAGCGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
QY 241 agggagacatgttaggtatcgagcacaacatccgcgactcgtgtggagagagagagag 300
Db 241 AGGGACAGTGTAGGTATCGGACACACATCCGATCCGATCGGTGGAGACGAGACTGCATGG 300
QY 301 gacaccccaacacctgagttctacagaaatgagatcgcttccttcgccaagcgctgttc 360
Db 301 GACACGCCCAACCTGAGTTCTACGAAATGAGATCCCTTCTGCGCCAAAGCGCTGTTC 360
QY 361 atgagagacatcttcagaaatcgagcacaactatgacctcctctgagagacatcctcc 420
Db 361 ATTGAGGACATCTTCAAGACTGAGCGGACCAACTATGACTCTCTGAGGACATCACTCC 420
QY 421 taactccagtgctgttccctcctcgagaaacacagagagtgagactggcatgccaagccctc 480
Db 421 TACATCCAGTGTGCTGTTCCTCTCGAGAACGAGAGTGGACATGCGACGACGCGCTTC 480
QY 481 acgctcagagagagtcgagagtggttaaaagctcccgagagatccagagcgctgttc 540
Db 481 ACGCTCAGAGAGGTCGAGAGTGTAAAGCTCCAGGAGATCCAGAGCGCGCTGTTC 540
QY 541 gctcagagagtcagtcgtgagcttcacgagatccgagtcgagagcgagagcgagcg 600
Db 541 GCTTACGAGCTCATGCTGCTGCTTACCGGAGTCCGGCTGAGAGACCGAGCGGCGACG 600
QY 601 gtgggc 660
Db 601 GTGGGCGGAGCAGAACTACCAAGAGCGCTTCCAGAACTGCGGACGCGACACAC 660
QY 661 aacctccgcacacacgc 720
Db 661 AACCTCCGCATCACACGATCTCAAGTCCGCTGTGAGCTGAGCTTGACACTTCCAG 720
QY 721 ggcgcacatgctcgccttccttcctgagagagagcgtgtgagcgcgagagctgcgaggtg 780
Db 721 GCGCAGCTGCTCCGCTCTCTCTGAGAGAGAGCGCTGTGCGGAGACTCCGCGGAGTG 780
QY 781 cgcgcagagtcgctcgc 840
Db 781 CCGGAGAGTGCCTGAGATCTTCAATGTTCCGCTGCTCCGACACACGCGCGCAG 840
QY 841 ctggtgcacatcgc 900
Db 841 CTGCTGCACTTCCCTGCGGACACTTCCGCGCGCTGCAAGTTCTGTGGGCGGCCAA 900
QY 901 gacacagctgcgagaggttcaaacccagctctctgcccacatccgcctcagagagagag 960
Db 901 GACACAGCTGCGGAGGTTCAACCCAGCTCTGCGCGCATCCGCTCGAGAGGCTCCAGAG 960
QY 961 gtgag 987
Db 961 GTGAGAGGAGAGAGACCTGACAGGGAGC 987

RESULT 9
AF156878 2208 bp mRNA ROD 16-JAN-2000
LOCUS AF156878 Rattus norvegicus Oploid growth factor receptor mRNA, complete cds.
DEFINITION AF156878
ACCESSION AF156878
VERSION AF156878.1 GI:6693831

KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2208)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Cloning, sequencing, expression and function of a cDNA encoding a
receptor for the oploid growth factor, [Met(5)]enkephalin
JOURNAL Brain Res. 849 (1-2), 147-154 (1999)
MEDLINE 20074706
REFERENCE 2 (bases 1 to 2208)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-1999) Neuroscience and Anatomy, Penn State
University College of Medicine, 500 University Drive, Hershey, PA
17033, USA

FEATURES
source
location/Qualifiers
1..2208
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/issue_type="brain"
/dev_stage="18 day-old fetus"
151..1893
/note="[met-5]-enkephalin binding protein"
/codon_start=1
/product="oploid growth factor receptor"
/protein_id="AF25002.1"
/db_xref="GI:6693831"
/translation="MDPDCDSWESESEDEGDGADDTTDETDGDDGDAEAPRS
LFQSRMTGKRMWPMQDMORYRHNYPDLTDQDNGMCLNIFKNEICFOPNALIED
ILQMKDNDYDLLENHSYIQWLPREPGVNMHAKPLTLEKVAEFSKREBRLVRA
YELMWFGEHLEDRGTGAVCAQNFQPRFHNHNSHNHNLRTTRILKSLGEGLYRA
QAPLVFEFLEETLVOKRLPVSQALDYFLFAVCRHORELYFPAWEHPRPREVW
GPRDKLRREKPORTIIPQILGPGQADDEGRDSORAGQRTGCGSGRSLGDSGTAE
DPSLINTKPEDGTLIDGNORDEKKSIPKSKRKRLKGNRODEVPEDADQGVSEYK
IALNLECALSPISQPREAEPPCPVAVANVEYRKRYEEGAEIGVSVNTQWQSA
LPTPSECEPAQSDGNGEDNSQVAEDSKSQVGEDPNSQVLEDPNSQVGEDPN
SQVGEDPNSQVGEDPNSQVGEDPNSQVGEDPNSQVGEDPNSQVGEDPNSQVGESE
ELRIEASAEPPKP"
1549..1764
/note="imperfect"
/rpt_type=direct
/rpt_unit=1549..1575
2031..2036
polya_signal
polya_site
BASE COUNT 522 a 615 c 675 g 396 t
ORIGIN

Query Match 30.5%; Score 698.4; DB 12; Length 2208;
Best Local Similarity 70.0%; Pred. No. 1,6e-93;
Matches 1046; Conservative 0; Mismatches 376; Indels 72; Gaps 5;

QY 48 ctgagcattccactcgtgag 107
Db 135 CTCTGCTTGTGCGAGTATGACGACCGGACTCGGATTCACCTCGGAGGAGAGCGA 194
QY 108 cgaagagcgagagagcgccgc 167
Db 195 GGAAGATGCGGAGATGCGCAGCGGATATACGACCGAGTACGAGGAGGAGGAGGAGTGA 254
QY 168 ggaagagcg 227
Db 255 CGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 314
QY 228 ctggcgagcacaagagagacatgltagatcgcgcacaactatccgagatcgtgtgagag 287
Db 315 CTGCGGTGCTATGACGAGCATGCAAGATACCGGACACCACTACCGGATTTGACAGATCA 374
QY 288 agactgcaatggggagacacgcacaaacctgagtttctacagaatgagatccgcttcctgcc 347

Db 375 AGACTGCATGTGATGATGCTGAACCTGAGCTTCTACAAAATATGATGCTTCCAGCC 434
Qy 348 caacgycgttttcaatagagacatcttcagaacttgagcagcaacatgacctctga 407
Db 435 AAATGGGGCTCTCTACGAGGACATTTCTTCAAGAACTGGAAAGACAACTATGACCTCTCGA 434
Qy 408 ggaacatctactccatccatccagtggtcttctctctctgagaaacagagtgatgca 467
Db 495 AGAATACTACTCTACATCCAGTGGCTGTTCTCTCTCGGGAACAGAGTGAAGTGGCA 554
Qy 468 tgcgaagcccccctcagctcagcagggagtggttttaaaagctcccaagagatccagga 527
Db 555 CGCCAAAGCCCTCCTACCTGAAAGAGGAGTGAAGCATTTAAAGCTCCAAAGAAAGTCAAGAA 614
Qy 528 ggcgtctgtccgagcctcagcagctatgcttggtcttctcagggatccggcttgagagcg 587
Db 615 GCGTCTTGCTCGGGCTATGAGCTCATGCTGGGCTTCTATGGGTTCCACCTTGAAGACCG 674
Qy 588 aggcagcggagcaggtgagcagcagcagaaataccagaacgcgttccagaactgaactg 647
Db 675 GGGCAAGGGTGTCTATGCTGACAGAACTTCCAGCCCGCTTCCACAAATCTGAACAAG 734
Qy 648 ggcgaagcacaacacacacctcgcacacacacacacacacacacacacacacacacacac 707
Db 735 CCAAGCCACAAACAACTGGGTATTACAGCATCTCAATCACTGGGTAGCTGGGCTT 794
Qy 708 cagagcaattccagcagcagcagctatgctcgtctctctcaggaagagcagctgctgagcg 767
Db 795 AGAATCTACACAGGACCCCTGGTCCGCTTCTCTGGAAGAACCTTGTATACAGCAAA 854
Qy 768 gctgcagcaggtgctgagcaggtgctgagcagcagcagcagcagcagcagcagcagc 827
Db 855 ACTGCCACAGGCTGGCAGAGTGCCTGAGCTACTTCTGCTGCTGCTGCTGCTGCTGCTG 914
Qy 828 cagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 887
Db 915 CCAAGCCCGGAGCTGTGACTTCTGCTGGAGCACTTCAAGCCCTGCGGAGAGTTGT 974
Qy 888 ctgaggg 947
Db 975 CTGG 1034
Qy 948 gggctcaggaaggtgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1007
Db 1035 GGGACCCAGGGGAGGAGATGAAGATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1094
Qy 1008 cagaggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1064
Db 1095 CCAAGGTCGAGCTGTGATCTGGAAGGAGCTGAGTGGGAGCAAGTGAAGCAAGCTGAGGA 1154
Qy 1065 gcccacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1124
Db 1155 TCCCTCAGTGTGAGAACAAAGCCCTCAATGGGGGAACCTTGGATGGGAACCAAGGGA 1214
Qy 1125 tgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1184
Db 1215 TGAAG-----CTAAGTCCCTGAGTCCCAAGGAGCAAGAA 1250
Qy 1185 gggagagcctgagagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 1244
Db 1251 AAGGAAGTTGGAGGGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1310
Qy 1245 ctcaagaggtgaggaagatcgctctgaatttgagaggggtgtgcccacacagcagcagcagc 1304
Db 1311 CTCTGAGGTAGAGAAATGTCCTTAACCTTGAAGGAGTGTGCTTGAAGCCCTATGAGC-- 1368
Qy 1305 cagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1364
Db 1369 -----CAGAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1400
Qy 1365 ccaacccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1424

Db 1401 T-----CTGGCCAGGGTGGCTAATAGGTAAAGAAAGCGAGAGAGGTGAGAGAG 1451
Qy 1425 tcttgaggagcagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1484
Db 1452 GGTGAGGGTGTATGAGTAGT-----CAGTAACTCTAAATGCAAGGAGGAGGAGGAGGAG 1505
Qy 1485 cccctgcccatcgaggcagcccaagcctgagcagcagcagcagcagcagcagcagcagcag 1538
Db 1506 TCCTAACCCCTTCAAGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1559

RESULT 10
HH092288/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Human herpesvirus 6.
Human herpesvirus 6
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Betaherpesvirinae; Roseolovirus.
1 (bases 1 to 30752)
Kosuge, H., Isegawa, Y. and Yamanishi, K.
Nucleotide sequence analysis of a 30-kilobase-pair region of human
herpesvirus-6B (HHV-6B) genome and strain-specific variations in
major immediate-early genes
Virus Res. 52 (1), 1-14 (1997)
98115230
2 (bases 1 to 30752)
Kosuge, H., Isegawa, Y. and Yamanishi, K.
Direct Submissions
Submitted (07-MAR-1997) Virology, Research Institute for Microbial
Diseases, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565, Japan
1. 30752
location/Qualifiers
1. 30752
/organism="Human herpesvirus 6"
/serotype="B"
/isolate="H57"
/db_xref="taxon:10368"
1. 1281
/gene="H74"
1. 1281
/gene="H74"
/function="replication, helicase-primase complex"
/note="part of helicase-primase complex, similar to HHV6A
U74, HCMV UL102, HVS gene 41, EBV BBLF3, EBV 54, VZV 52,
HSV1 UL8"
/codon_start=1
/product="H74"
/protein_id="AAC40324.1"
/db_xref="GI:2769712"
/translation="LOSVAFLERHIFGVYKOVOLVCEYIGPDLMPGNGERSVFIG
YPMILMISVSDLEKRPDTTYICRELLISFGAPLIGRRAIPVIRELSVEMGSE
TSLORFNSOVSSESLCFORGPDTLFFESDSMVTYLPDCIRLLKSTVPAPL
PCDENATRELLIKEMSLQRSALDVAIVTMDAVSAFORACTLMEMKLVLRD
LHYFTCTGCKRSHVYMLDTVAENCKEITELKORPAFCOMETSRGGEVFAAEFL
SSGSLSKAHMTVATSKYLISACIRANVTGLCFASITVFDMMCVETANRYNYSWI
EEFDPNDYCLEHGLDCSRATVAMSDDGLVROARSIATLDKINFSYTLVTLRVL
RMVWSEKPEVDVQTEFIEMENRMALYEHILHLIN"
/complement(1316..2065)
/gene="H75"
complement(1316..2065)
/gene="H75"
/note="similar to HHV6A U75, HCMV UL103, HVS gene 42, EBV
BBLF2, EBV 55, VZV 53, HSV1 UL7"
/codon_start=1
/product="H75"
/protein_id="AAC40325.1"
/db_xref="GI:2769713"
/translation="WKSILKLEKLETSDFNTLVHRTILKVIKIDKCVSLARHSLVNI
TVGDGDIWFHLEDGTIMNGLEKYKICEKELGFGFGIILIDSBDTLQELRLNPFQK

gene
CDS
/gene="H76"
/function="possible virion protein"
/note="similar to HHV6a U76, HCMV UL104, HVS gene 43, EBV BRF1, EBV1 56, VZV 54, HSV1 UL6"
/codon_start=1
/product="H76"
/protein_id="AAC40326.1"
/db_xref="GI:2769714"
/translation="MHRASANSLNLSVSGSMRRNOSGRRPSKRLSDNENATLSTNS
ILGADMLSKNLSTYLPNNNEEDIMYPSQIQLFTIMLGHKNFRGQITTHALRDS
AVLKQDAIVAAQALLNSVSIQIIDHEKRWKHSVSPHNKLSFDYFSWAIAIKQV
ILGDISINIFIIQSIDNSHNYVMICVGVPEPRTPTAPNLNLYNQVSSKLI
HDIYRKONIVITPILGLSSVILPDEFINIKIFPDNRNSEQISCFKNKKAIAFYSPY
VIRNRMLTTPPLAHSPELKNHSLRRHOCQOLNTEPFIKVLITTAQTDVTKNDM
IEEKNNSDAKSLIKELNLSDSKISQIGTIDSVGEIOETPSIIDONLMDRGOF
RKRSALDTGREDVDLFRKQIYKCMESQITQDDEFTLKTQNMERKIKDHSLE
TNDCORNPDLHDLENLSLRALNIVQLPPTSVIDTDSVANSFESQIYPTQY
ADKRIDQLMEWEMKRTFRLKRNVNQOQESITYSNSELIVPFLRRLNLTNYES
IPEEFLSLIGEILAIYESSKIKHYLRVYRELNOISEVYMLTQTHPENNPIDFS
NIRSPNENEILKIKIRINLRIRLOHTRPNPKGDQD"
3757.6231
/gene="H77"
/gene="H77"
3757.6231
/gene="H77"
/function="helicase, helicase-primase complex"
/note="similar to HHV6a U77, HCMV UL105, HVS gene 44, EBV BRF4, EBV1 57, VZV 55, HSV1 UL5"
/codon_start=1
/product="H77"
/protein_id="AAC40327.1"
/db_xref="GI:2769715"
/translation="MSISLFGGRYDKNFKLLNMSAPKIELIVDKVASISERLEGTL
PEMDRHIMDETEFNSERPADACIGIDEPQPLPPLFRALLVYTGACAGKNTSOT
LANLDCITVATTSIAQNLNVNLSKSAQVTKIRTFEGSSHVMSRQSYIAND
ERSIQIQKQDLSIYNNVLSIDIAERALGAACFKELPDCESSVPIVIEAGVILRI
LHTVFEFYENALYKLPYLENGIVPCIVGSPQSNALYTSFNPLOKDKVGRGD
VLSALICDVLKSYCEVDNMILFVNKKRCADAFGDELKIEFGLEKPELLEYDO
FVKPASYIRNPMNEIETTRFLSHNEKKNRFSRHQVETNNNNLEVPVYFLINK
TEEDYKSEIGNESELEIEPMKSNINHNNTSQPADDLSTVQLEELVLEDGVEELT
ITCHLHNRNSIGVSKITASTVFGSGTEKVELQSDLEKTSCEVTIAYISFL
SGLFGMYSFSCSEFTTPEVMEIKNIMKPSIEFSEKSRSRSDVQVETDERDF
GLVDDGLSDDLLEIDPCGDPFTFYSKLP/LNLSFESEISLTYTEKDFISRFAL
OMHTKGFGKTLVYNNRNNSRKGGEISYSLKSFQMLTYAIPANNYTLLEGYTDN
VYHLGDKOLPOLIYKKGLPRIYIKDEMGRISVLDDNNSKFEUVVNGSQSHLCTTYDY
ATYKSYMTITTKSQGLSIQKVAIDFGSDPKNLKLSTIYVGMKSVTDPNNLIMVNPDLR
LNTENDFIAPHIVKALKNENTMLIF"
6495.6722
/gene="CB5RH"
/gene="CB5RH"
6495.6722
/gene="CB5RH"
/codon_start=1
/product="CB5RH"
/protein_id="AAC40342.1"
/db_xref="GI:2769730"
/translation="MTSAKTNLSLPSTRTMYSGNLKNQAFSLSSRELNVNFDVNS
GGGLDGLVSTNRSSTIECTTEVILGIGVRLR"
8179.8406
/gene="CB6RH"
/gene="CB6RH"
8179.8406
/gene="CB6RH"
/codon_start=1
/product="CB6RH"
/protein_id="AAC40343.1"
/db_xref="GI:2769731"
/translation="MLHNSKCIHSICIRVCILHWYCSSRFSIRGFAVETYSILPN
TSDGQAFLYAGFKVHALVADAIRKGFRRG"
8829.9866
gene

CDS
/gene="H79"
8829.9866
/gene="H79"
/function="possible replication protein"
/note="similar to HHV6a U79, HCMV UL112, spliced in vitro replication protein"
/codon_start=1
/product="H79"
/protein_id="AAC40328.1"
/db_xref="GI:2769716"
/translation="MTAERGGYGSFDNVIAQVEQIISQSLHKKREFDNGCFIEFLAD
SGYCEITSGKWSIMYIWTSEITDSMGLTYDIDGDEGCRITYARGLLCSKITSISQ
NEGGERILTVSHENGKLIQITVYTIKVSSEPLRNIGDLKFEKFEKCRALDRKKH
DDEHRRSGKOKREKREVEDIDKKDEDEKIKOEKKNDEKDPKDEDFDGKYECAV
VNEFTIGYRSRTFLNSICFPVYISEPKRKSHHETRNLEBOSEHGCIAPTSTP
VNGAVGALSPCVSIDNHHDDQDHELDKRVIAQVGSVLSPPKRSLESLICVSKADLF
LIGDEPFRSIO"
9838.10449
/gene="H80"
9838.10449
/gene="H80"
/gene="H80"
/function="possible replication protein"
/note="start-codon ggg; similar to HHV6a U80, HCMV UL113, spliced in vitro replication protein"
/codon_start=1
/transl_except="(pos:9838..9840,aa:Met)
/product="H80"
/protein_id="AAC40329.2"
/db_xref="GI:7687902"
/translation="MTNREGLSMEDYNGNFOLNKLSLEQRLARLVASGGEVERSLSKL
KERLDYKQNLKRVNLECAQVYPSKLSKTRHIEQKQIVSDCVSYVCEIKFEI
DMRVETERTQARVRYORRTYVSTGDAICQQRISQONRGRAGRGRVRRRN
SNLNIRTONSAIVIDDSETERENAGSFNEDLATYTLTTL"
complement(10493..11260)
/gene="H81"
complement(10493..11260)
/gene="H81"
/note="similar to HHV6a U81, HCMV UL114, HVS gene 46, EBV BRF3, EBV1 61, VZV 59, HSV1 UL2"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="AAC40330.1"
/db_xref="GI:2769718"
/translation="MALLQWMLDHODEKKNYENISIDPHSILFGINRWLSFLOLST
LEITHLKHVYKLVNDRAHLYVHPSDDNHNANSFLCKPVDYVYIIGHDPYDGKCG
LAEVYKQESISESLKINIKERLESIPNPSPDNGCLNCKEGLVLLNSITVYHGL
PMSHEAFGQATSYKIIISRLSEQMSLVILMGKHKARKSYLIDAKHLVLEBSAPHS
KYAAMPPIGCHHEFRTNLFLEHGKDPINNNILNE"
complement(11336..12088)
/gene="H82"
complement(11336..12088)
/gene="H82"
/function="gH-gL complex, infection, cell fusion"
/note="gH; similar to HHV6a U82, HCMV UL115, HVS gene 47, EBV BRF2, EBV1 62, VZV 60, HSV1 UL1"
/codon_start=1
/product="glycoprotein L"
/protein_id="AAC40331.1"
/db_xref="GI:2769719"
/translation="MELLFVMSLILKPSKAMPLEHDNHSFYEKLDCCIAAVINCTR
SEVPLILEPTIYQAPVYNEVMSLILKPKKPEFSIMYITNLSPLIDCINPEQLR
TIFALIGDEPSRDNMLNFPNGFOTGSPSVGITTICISDNCRKYLPEITTYVNNFVDNI
AGLEFNISNTDSFYSNIGFLYLENPAIGITRIKRPNSLTFEOTLILNCLKYEFLK
TGVFEDLLQMEAYNSKLPFRSSRPILIRNT"
12057.12506
/gene="H83"
12057.12506
/gene="H83"
12057.12506
Query Match 4.6%; Score 106.4; DB 72; length 30752;
Best Local Similarity 55.4%; Pired. No. 2.5e-07;
Matches 206; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
QY 1572 taccctcggagagcccatcgagagccagagccagcgagactcgagggagcga 1631

[illegible][illegible]

RESULT	11
AB021506/c	
LOCUS	AB021506 161573 bp DNA
DEFINITION	Human herpesvirus-6 (HHV-6), variant:B, strain:HST, virion genomic DNA, complete sequence.
ACCESSION	AB021506
VERSION	AB021506.1
KEYWORDS	GI:4995977
SOURCE	Human herpesvirus 6 (strain:HST, pop_variant:B) DNA.
ORGANISM	Human herpesvirus 6

[illegible]

REFERENCE AUTHORS	TITLE
1 (sites)	
Isegawa, Y., Mukai, T., Nakano, K., Kagawa, M., Chen, J., Mori, Y., Sunagawa, T., Kawamishi, K., Sashihara, J., Hata, A., Zou, P., Kosuge, H. and Yamanishi, K.	Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B
JOURNAL U. VIROL.	73 (10), 8053-8063 (1999)
MEDLINE REFERENCE	99412319
	2 (bases 1 to 161573)

```

//note = 82.3% identical to UK3 of strain UI102 of HHV-6"
//codon_start=1
//protein_id="BAA78214.1"
//db_xref="GI:4935981"
//translation="MSRVFSCVLRYACVCAGLICQCMVCMGIVICGDCQHMWRRCARNGRV
GPRLVADGAMVRSRGDGVRSYRSTCKTERASARAPYSPFVLYSSSSSSAC
SSRPSPSPSSHAASHAVCAEGGRDLPIMHADCDADGEGDITLKGADGADGAGNAT
GCPTDTPFAFSPGCDLGMXNAGDLGDEGEGTGGEGNDGGE"
//complement(5023..5532)
//gene="DRH1"
gene

```

FINAL
JOURNAL
DIRECT SUBMISSION
Submitted (18-DDC-1998) to the DDBJ/EMBL/GenBank databases. Yuj1
Isegawa, Osaka University Medical School, Department of
Microbiology; Yamada-oka 2-2, Suita, Osaka 565-0871, Japan
(E-mail: isegawa@micro.med.osaka-u.ac.jp, Tel: 81-6-879-3333,
Fax: 81-6-879-3329)

```

/gene="DRHN1"
/note="ORF unique to HST strain"
/codon_start=1
/protein_id="BAA78215.1"
/db_xref="GI:4995982"

```

FEATURES	Location/Qualifiers
source	1..161573 /organism="Human herpesvirus 6" /virlon /strain="HST" /db_xref="taxon:10368" /pop_variant="B" complement(18..365) /gene="I1" complement(18..365)
gene	
CDS	

gene	5025..5336	/gene="Dr6"	
CDS	5025..5336	/gene="Dr6"	
	/note="84.5% identical to Dr6 gene of strain U102 of H11-6"		

```

/gene="L1"
/ncbi="52.58" identical to L1 gene of strain U102 of
HIV-6"
/codon_start=1
/protein_id="BAA78211.1"
/db_xref="GI:4995978"
/translation="MVRARVRVAVSAMVRAMVRVEMVRAMVRVAVRAMVRVRVRY

```

```

gene
6512..7150
/gene="fba7"
/codon_start="1"
/protein_id="BAA78216.1"
/db_xref="GI:4935983"
/translation="MTTRTOMRDGRIARPDGARLAHARAREPMLILGRPSK
LYGSTRSRGSRIRHLPMRPWCLHDPDPRDARSATWGHKMGPPTHVPRRSYQDCG
E"

```

CDS 6512..7150
 /gene="DR7"
 /function="transforming activity"
 /note="42.2% identical to DR7 gene of strain U1102 of
 HHV-6
 US22 gene family"
 /codon_start=1
 /product="transactivator"
 /protein_id="BAA78217.1"
 /db_xref="GI:4995984"
 /translation="MRAARCRLASNTVEFARGLQBDPVAEEDLEPDRMCGTNV
 RLPEVIAADEHDLTVGGICOTHAQVSELEETVADPWTAVGVMTLTVAADY
 RLMPARARQLRHETAGSPGLSPWYICANLSNREQTQPSPIGCGYTLGNVPFPPGR
 EVELMAVITMRDRCISGPLTARSAVWRKNSPPGGWAGHETLDFGFG"
 complement(7236..7706)
 /gene="DRHN2"
 complement(7236..7706)
 /note="ORF unique to HST strain"
 /codon_start=1
 /protein_id="BAA78218.1"
 /db_xref="GI:4995985"
 /translation="MLAPPTPGVYIKGCGALSSPPLGCDSPCGVIGRCYSNAG
 RCLDRKYVPEFLGCVNTRKHVYCDVSRPLFLGLAGSDYTRTYHTPHRR
 RRIIDSTAVLEFLRLRKTPTFFFLVYFELVMAIFLRSCIMPTASRLRGV"
 7928..8662
 /gene="DR8"
 7928..8662
 /gene="DR8"
 /note="13.6% identical to DR8 gene of strain U1102 of
 HHV-6"
 /codon_start=1
 /protein_id="BAA78219.1"
 /db_xref="GI:4995986"
 /translation="MROVTRRDQKGRLLAWOLGKKNRCKILTKETSLHPLTPST
 TTPRLAHNFCGRFSRRVYOTQPTPHATHTMAVAGCAGISRGADYKRNGLAPRV
 PSSRLALARQGRPLNOSTITLTLTLTLTLTLTLTLTLTLTLTLTLTLTLTLTL
 LTL
 CAPRGSPLHPPPKNGYGGGWV"
 complement(8292..8807)
 /gene="LJ1"
 complement(8292..8807)
 /gene="LJ1"
 /note="16.6% identical to LJ1 gene of strain U1102 of
 HHV-6"
 /codon_start=1
 /protein_id="BAA78220.1"
 /db_xref="GI:4995987"
 /translation="MALNAVLSFLRLRLVCSAALPSASLSLSPAFESFNSHRAPAG
 CTFAKLAHPHTLARSIGGGGIGIPAGRTGDTGRGLGVRGREFHAPLAAVAYL
 VNTPAPLEAARALGDGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLG
 LGLGLGLGLGL"
 8929..9384
 /gene="U1"
 8929..9384
 /gene="U1"
 /note="25.6% identical to U1 gene of strain U1102 of
 HHV-6"
 /codon_start=1
 /protein_id="BAA78221.1"
 /db_xref="GI:4995988"
 /translation="MRAAPRAIDLSTDPRISVQJINTRTITDMPTRPHRADCVSE
 SHTEKHETKRYTOENTHQTOKYTHKIHQTKIHQTKIHQTKIHQTKIHQTKIHQTK
 HKTKYTHKKVYTHKTYHKKIHTHKIHTHKIHTHKIHTKNHPIYALFSGC"
 complement(9467..10768)
 /gene="U2"
 complement(9467..10768)
 /note="92% identical to U2 of strain U1102 of HHV-6
 US22 gene family"
 /codon_start=1
 /protein_id="BAA78222.1"
 /db_xref="GI:4995989"

[illegible]

1

[illegible]

BASE COUNT	5136 a	3043 c	2213 g	3572 t	1 others
ORIGIN					
Query Match	4.0%;	Score 92.2;	DB 72;	Length 13965;	
Best Local Similarity	52.8%;	Pred. 4.3e-05;			
Matches 199;	Conservative 0;	Mismatches 178;	Indels 0;	Gaps 0;	
QY 1561	cccaagaagtgatcccttgggaagcccatcggagaccccaagccccaagccagcagaacct 1620				
Db 9881	CCAGCAGAGGACCTCCAGCAGGGATTCTTACAGAGCCTCCAGTAGAGCCTCCAGCAGGGGCT 9940				
QY 1621	gcaaggggacgaagccagccagcagaagcccatcggagaccccaagcccccgcgcggcagaacct 1680				
Db 9941	CAAGCAAAAGCCTCCAGCAGAGCCTCCAGCAGAAGCCTCCAGCAGAGCCTCCAGCAGGGATT 10000				
QY 1681	gcaaggggacgaagccagccagcagaagcccatcggagaccccaagccccaagcccgccagaacct 1740				
Db 10001	CTAGCAGAGGCTCCAGCAAAAGCCTCCAGCAGAGCCTCCAGCAGAGGATTCTAGCAGAGCCT 10060				
QY 1741	acaaggatgagccagcccggaagcccatcggagaccccaagcccccgcgcgcggcagaacct 1800				
Db 10061	CCAGTAGAGCCTCCAGCAGAGGCTCCAGCAAAAGCCTCCAGCAGAGCCTCCAGCAGAGCCT 10120				
QY 1801	gcaaggggacgaagccagccagcagaagcccatcggagaccccaagcccccgcgcgcggcagaacct 1860				
Db 10121	CCAGCAGAGCCTCCAGCAGAGGATTCTAGCAGAGGCTCCAGCAAAAGCCTCCAGCAGAGCCT 10180				
QY 1861	gcaaggggacgaagccagccagcagaagcccatcggagaccccaagcccccgcgcgcggcagaacct 1920				
Db 10181	CCAGCAGAGGATTCTAGCAGAGCCTCCAGCAGAGGATTCTAGCAGAGGATTCTAGCAGAGGCT 10240				
QY 1921	acaaggatgagccagcc 1937				
Db 10241	CCAGCAGAGGCTCCAGC 10257				
RESULT 15					
HHUJ13194/c					
LOCUS					
DEFINITION	HHUJ13194 21191 bp DNA VRL 17-JUN-1995				
	Human herpesvirus 6 replication origin-binding protein (HHRF0), partial cds; helicase-primase component (HHRF1), virion protein (HDLRF1), putative helicase (HDLRF2), putative phosphoprotein (EDLRF1), replication protein (EDLRF2), putative uracil-DNA glycosylase (EDLRF4), putative 9L glycoprotein (BDLRF3), putative glycoprotein (BDLRF2) genes, complete cds, and major immediate-early protein 2 (BDLRF0) gene, partial cds.				
ACCESSION	UJ3194				
VERSION	UJ3194.1				
KEYWORDS	GI:862477				
SOURCE	Human herpesvirus 6.				
ORGANISM	Human herpesvirus 6				
	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae; Roseolovirus.				
REFERENCE	1 (bases 1 to 21191)				
AUTHORS	Nicholas,J.				
TITLE	Nucleotide sequence analysis of a 21-kbp region of the genome of human herpesvirus-6 containing homologues of human cytomegalovirus major immediate-early and replication genes				
JOURNAL	Virology 204 (2), 738-750 (1994)				
MEDLINE	95027704				
REFERENCE	2 (bases 1 to 21191)				
AUTHORS	Nicholas,J.				
TITLE	Direct Submision				
JOURNAL	Submitted (09-AUG-1994) John Nicholas, Johns Hopkins Oncology Center, Johns Hopkins University, 418 North Bond Street, Baltimore.				

MD 21231, USA
On Jun 17, 1995 this sequence version replaced gi:662095.
COMMENT
FEATURES
source

Location/Qualifiers
1. .21191
/organism="Human herpesvirus 6"
/strain="U1102"
/specific_host="Homo sapiens"
/db_xref="taxon:10368"
/clone="HindIII-D, EcoRI-D, pPCR-3"
/map="109-130 kbp"
1. .1980
/gene="HDFR0"

gene

CDS
/gene="HDFR0"
<1. .1980
/codon_start=1
/product="replication origin-binding protein"
/protein_id="AA68464.1"
/db_xref="GI:662096"

translation="SARLTENVDLILDEIMSIKKPFYKMTKKEVDCKFLSLIK
NSSHYIAMDATLTHRVVEFAAFKPDQIALIRMTFVSAMFSNRIAYRCDFPKKEFS
FPALDEKTRMDKLCLECSVLAETMYDLIRSFPAKILLTSKOGKSSIESNI
RYDVITVSVTGLSEFPVFSLEFVIQLAKGPDVSVQSIGRVKRVIDEITLI
YMPVLINSYDPLAIPCDSDMSVAOSIISSECIDFRKCSGAHKNVCVSLK
FRYRIYEKTTISLSLFLCISLACENSITKDVIVGNFPMKREVFSLFQIIVEEC
HFIERKITLPDMDTFOEIISSRETIMGDEYENDQDLHRDYITDMKFPATFISPG
VDIFIASDIIVYDIKNEKRVYVNWVLOKVSAGVSETEIRVFCERIKCYMLKRSFL
CDEYFVGLDISGYEMKMLIDLAFLAEIRKDKLKSCTDTTDDISEDLLICARRS
SDIIQIMOLVETVHVFQKYSLOTQLFNLRKGRVITGVFSIEKFSISILRLEFK
AFNMTLKSDRPRIIPGKAYRNLTAKMDMENMDNMEISTNTLKTCKELKALTEASRAR
RKQIILYKLGSDISLSVSEVGFQGHASPGVCS9"

gene

CDS
/gene="HDFR1"
1949. .3937
/note="Component of helicase-primase complex; similar to
HCMV UL102 Swiss-Prot Accession Number P16827"

/codon_start=1
/product="unknown"
/protein_id="AA68465.1"
/db_xref="GI:662097"

translation="MHLRGCAHSLYCVYNDWENKIYVPIFQCLFLEAETRSIKTF
LIRGSLDOESLINEIEYRKETMLMDLOESNMMDKIAAISLIMNGELVRLSKF
FPLPLTVVGGDGLLEIYVAGEELMLPDTVPVLCIGDYALDTHLSNBDTOA
SKTIRFGYSVMDPFLSLVSDVKICFSTHTDTPGAECMKROIIFSVCARFGEDVILY
TPKNAYALLFPDDCCLILLOSVPAPFLHEKIRIGYKQVIVOCETIGIPMLPGKRSY
SFIGPMLWLSVSDLERKVPDTYITCEIILSFCGLAIPDRGHAVPVARELSIEM
PGSETSLQRFERNOSYSESILCFQOTGEDHLEFSDDMVTVLPDCLRLLKSTVP
KAPLPCEDENAETIDLLKFMRSYALFDVAVIFMLDAFVSAFORACILMGKRMV
LVRLDHMYLTCGDGKDHVMPLOTAVENCMKETEIKORFQACASISGCFVVA
RFLPSGSLGSKSEAHMYTASRYSACIRTKTGICFASISIVYFODMMKCVIARVYV
SVNIEEPNDYCLEYHEGLDSCSTYTVKSEDDQVYVQARGIALITDKINFTYILVT
LVRLRWVESKFEDVEQYTRWENRMLCEHILHLHN"

polyA_signal

gene

polyA_signal

gene

polyA_signal

gene

CDS

/note="similar to HH6 CBZL gene product, GenBank Accession
/gene="HDLF2"
/evidence-not_experimental
complement(3972. .4721)
/gene="HDLF2"
/evidence-not_experimental
complement(3939. .3944)
/gene="HDLF2"
/evidence-not_experimental
complement(3939. .6618)
/gene="HDLF1"
/evidence-not_experimental
complement(3939. .3944)
/gene="HDLF2"
/evidence-not_experimental
complement(3972. .4721)
/note="similar to HH6 CBZL gene product, GenBank Accession

CDS

Number L14772"
/codon_start=1
/product="unknown"
/protein_id="AA68466.1"
/db_xref="GI:662098"
/translation="MKSIRKLEKETSDVENILHVRITLKVIKIDKICISLARPLVNI
TVGDGIMFHELDGTMINGLVEYKICEKEGFOGFIITLDSHDTLOEIRLNPFOK
RRIJHMKVDPEEPMGLVFALENLEKOSTLHKLIAKVLVLPALSPITKILFNCD
TLVOTLARIHFNPHASILKRPVIMITLYNEMKNTHEVLELFTNKRSHNFINLSLE
SRQJDSLSLOYIQTATQFAQTFYSKNGDTS9"
complement(4630. .6618)
/note="HDLF1"
/gene="HDLF1"
/note="similar to HCMV UL104, Swiss-Prot Accession Number
P16735"

/codon_start=1
/product="vtrion protein"
/protein_id="AA68467.1"
/db_xref="GI:662099"
/translation="MHRASANSLSVSGSMWRNROSSGRRPSKRLSDNATLSTINS
ITGAEMLSKNLSTSLPNNNEIDMTPSPQIMPTFIMLGHKNFEFGOTIHNAIRDS
AVLKQIAYGAQALLNSVSIQIHDKMKHNSFPFNKKLSRODYFWAHAIKQY
ILGDISINIFITQSIDNSHNRVMDICVGVIPERRTHLAVNLNLOOVSSKI
HDIVRAHQNTVTPVLGLSSVYIPDFNRIKIFRDNSEQISCFKNKAIAFPYSDPY
VIRNRMLTTPPLAISPCLKHNSLRHROKCOLANFPPIKVLITTAKTDTYNNKIMDM
IEKEKSDAKSLIKRFLNLSDSKSIKIGRDSVEGFOETTPSIIDONKMLNRGOF
RKSAIDTGERVDRLFKKOLICMEQIOTOMDEITETITTONMFERKIKIDLSLE
TNDCDNYNPMULDELNLSRALNIVQRLPTSVSIDOTRSVANSFQOYIPDTY
ADKRIDOLMEYETRTFLRNVNNOQESITYSNISILLVPLRLRLNLTYNLES
IPEEFLLSGEIILAYESSIKIRIYLVYVRELNOISEVFNLTQHPNPSPIFDS
NIFSPNENELIEKIRIRIHLRIRIQLHTRPNYPKGDQ9"

gene

CDS

6413. .9409
/gene="HDFR2"
6413. .8887
/note="putative helicase; similar to HCMV UL105,
Swiss-Prot Accession Number P16736"

/codon_start=1
/protein_id="AA68468.1"
/db_xref="GI:662100"
/translation="MSTISLLEGGRYDNKFLINSSAPKIELIVDKVASLSERRLEGR
PEDFRIIMPETEFEFNSPEADALCIGIDEPAOLPLPFPALAVTGAGAGKINSIOT
LAANLDCIVATATSIQAONLSVYNRKSQAVKTIFPTFNSHVSMSROSYIAND
ERSIIOQKODLSITVMVDSIDIRALGAVACKRELPLCESSVIYIDKQVYLRHI
LHTVPEYKMYNALKRPIKLEDGIVYRIVCGSPQSNALVTSNPLTQKDKVGRGD
VLSMLIDVDVLSKICEVDNMWITFVNNKRCADHAFGFLAHIEGCLPLAKELLIEYDQ
TFEDPKSEIGNEIETRLFLSHNEVKNYFRSLHEQVETNRNNLFEFVYPLINK
ITCLHRIKINSISIGVSKIRASVYFGIYERKEVELLOSPLIEKTSQDOTIAHVSFL
SGLMEGMYSPCCSKFTTPVYMLIKIKMPSIEFLSEMSRMSPDVQVETDERBYE
GLVNDGISDVLLEIDPCGDPFETRYKELLTNSLSPEELSIATYTKDILFIRFALL
OKHTKGFQGTLLVTVYRNNVSKROCEETISHLKSFQMLTYAIPTANPTLBECTDN
VHLHGTQKQPLQILYKRGPLVLRKDMGKISVLDNNVSKEFIDVNNQSHILCTTVY
ATVSKVSMITITKSGLSIQKVAIDFGSDPNLKLKSIYVGMRSVTDENNLMNNVPLR
LNYENDNFIAPHYKALKNENTMLIF"

polyA_signal

gene

polyA_signal

CDS

gene

polyA_signal

gene

polyA_signal

gene

CDS

/note="similar to HH6 CBZL gene product, GenBank Accession
/gene="HDLF2"
/evidence-not_experimental
complement(3972. .4721)
/gene="HDLF2"
/evidence-not_experimental
complement(3939. .3944)
/gene="HDLF2"
/evidence-not_experimental
complement(3939. .6618)
/gene="HDLF1"
/evidence-not_experimental
complement(3939. .3944)
/gene="HDLF2"
/evidence-not_experimental
complement(3972. .4721)
/note="similar to HH6 CBZL gene product, GenBank Accession

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2000, 01:35:17 ; Search time 1481.96 Seconds
(without alignments)
9387.105 Million cell updates/sec

Title: US-09-431-843B-1
Perfect score: 2250
Sequence: 1 tgggctcagccagcccccag.....aaaaaaaaaaaaaaaaaaaaa 2250

Scoring table: IDENTITY_NJC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estp16:*
76: em_estp17:*
77: em_estp18:*
78: em_estp19:*
79: em_estp20:*
80: em_estp21:*
81: em_estp22:*
82: em_estp23:*
83: em_estp24:*
84: em_estp25:*
85: em_estp26:*
86: em_estp27:*
87: em_estp28:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss15:*
97: gb_gss16:*
98: gb_gss17:*
99: gb_gss18:*
100: gb_gss19:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: em_gss12:*
109: em_gss13:*
110: gb_gss5:*
111: gb_gss6:*
112: gb_gss7:*
113: gb_gss8:*
114: gb_gss9:*
115: gb_gss10:*
116: gb_gss11:*

|||||
Db 366 AGGACACAGAGTGAATGCGATCCAAAGCCCTCACACTGAAGAGGTGAGCATTT 425
Qy 532 aaaaatccaaagaaagtaagaagcgtcttgcgggacctatgatctggtcttc 651
Db 426 AAAAGCTCAAGGAAGTGAAGAGCGCTTCTCCGGGCTTATGACCTATGCTGGGCTTC 485
Qy 652 tatgggtccacctgaggaacggcggtgtctgtatgcgtgcacagaactccag 711
Db 486 TATGGGATCCAACTTGAGAGACCGGAACACAGCGCCGTATGCGGTGCACAGAACTTCAG 545
Qy 712 ccgagctcccaaatctgaacagcacagccacaacacctgagttatcacgccttc 771
Db 546 CCAGCTTACACAACTCAAAAAGCCACAGACACAACATCTGCTATTACAGCATCTCTC 605
Qy 772 aagtcacgtggtgagctggtgctagaac 799
Db 606 AAGTCGCTGGCGAGCTGGGTTAAGCAC 633

RESULT 2
A1115047 550 bp mRNA EST 02-SEP-1998
LOCUS A1115047
DEFINITION u41f12.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885007 5', mRNA sequence.
ACCESSION A1115047
VERSION A1115047.1 GI:3515371
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:969331

FEATURES
SOURCE
Seq primer: custom primer used
High quality sequence stop: 531.
Location/Qualifiers

1..590
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_image="IMAGE:1885007"
/clone_id="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACGCTG);
Site_2: DraIII (CACCATG); 1st strand cDNA was primed
with an oligo(dT) primer (ATGCGCTTTTCTTTTCTTTT);
double-stranded cDNA was ligated to a DraIII adaptor
(TGTGGCCCTACTGG), digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACGCTG, 3' site
CACCATG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAGCTGGC and 3' end primer

BASE COUNT 160 a 152 c 173 g 104 t 1 others
ORIGIN CGACTGACGCTCGAGCACA."

Query Match 24.1%; Score 542; DB 8; Length 590;
Best Local Similarity 95.5%; Pred. No. 3.6e-109;
Matches 557; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 136 tccgtgtctgagatgagacagaccgagctcgattccacctggaggaggagcgag 195
Db 8 TCCAGCCTGTGAGATATGAGACGACCCGGAATGCATTCACCTGGAGAGAGAGCGAG 67
Qy 196 gaggatggcgaggtgtggcgaagcgatgatcgccagatgagcagcgagcgagatgac 255
Db 68 GAGGACGGCGAGGATGGCCAGCGGATGATGACCATGAGTGGACGGCGAGGATGAC 127
Qy 256 ggcagcgcgagagagagcagcgaagcctgttcagtcacagatgacaggtaccgaac 315
Db 128 GCGGACGGCGAGGAGGAGCGGCGCAAGCCTGTTCAGTCTAGATGACAGGTACCGAAC 187
Qy 316 tggcgtgctatgcagagacatgcaaaagabaccgagcaactaccgagattgacagatcaa 375
Db 188 TGGCGTGTATGACAGGACATCAAAAGATACCGGCAACACTACCGATTGTGACAGATCAA 247
Qy 376 gactgcgaatgtgacatgtgtaacatgagcttcaaaaaatgagatgctgttcacgca 435
Db 248 GACTGCAAGGAGACATGTGACACTGTGCTTACAAAATGAGATCTGTTCCACGCA 307
Qy 436 aatggggtctcctcagaggaacattcttcgaagactggaagaaacactatgacctctgaa 495
Db 308 AATGGCTTTCATTTAGAGAGCATCTTTCAGACAGGAAAGCAACTATGACCTCTGGAA 367
Qy 496 gagaatccttcacatcagtggtgttctctctgcggagaaacagagatgaactggcac 555
Db 368 GAGATCACTCTCAATCAATCAATGCTGTTCCTCTGAGGAGAACAGAGTGAATGCGCAT 427
Qy 556 gccaaagccctcaccctcgaagagaggttgagcatttaaaagctcgaaggaagtaagag 615
Db 428 GCCAAGCCCTCACACTGAAGAGGTTGAGGCAATTTAAACCTCAAGGAAGTACAGAG 487
Qy 616 cgtctgtccggcctatgagcattcatgctggtgcttcatagtgtccacctgagagcg 675
Db 488 CGTCTGTGCGGGCGCTATGAGACTCATGCTGCTTCTATGGAGATCAACTTGAAGACCG 547
Qy 676 ggcaggggtgctgtatgcgtgcagagaactccaagccgcgt 718
Db 548 AACACAGGCGCGGTATGCGGTGACAGACAGAACTTCCAGCCACGCT 590

RESULT 3
BE534864 735 bp mRNA EST 09-AUG-2000
LOCUS BE534864
DEFINITION 601231408F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3595608 5',
mRNA sequence.
ACCESSION BE534864
VERSION BE534864.1 GI:9763509
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 735)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.

OY	316	tgggctgtcatctcagaagactgtcaaaataacccgcaaacactaccggatttggcaagtca	375
Db	187	tggcgtgcattatgcaggacactgccaaagatgccggcaactaccggatttgcagatcaa	246
OY	376	gactgcaatlygtgacatlygtgcaacctgaagcttctacaaaatlyagaatcgtctccagcca	435
Db	247	gactgcaaaaggagacatgtgcacactgagacttctacaaaatlyagaatcgtctccagcca	306
OY	436	aatggggtctctatcgaagacattcttcaagaactggaaagacaactatgaactctctgaa	495
Db	307	aatgcttctctatgagagacattcttcaagaactggaaagacaactatgaactctctgaa	366
OY	496	gagaatcaactctacatccgaagtcgtttccctctcggagaaccagagttgaacttggac	555
Db	367	gagaaatcaactcttactacatccgaagtcgtttccctctcggagaaccagagttgaacttggac	426
OY	556	gccaaagccctacacctgaagaggtltgagcatltaaaagctccaaagaaagtcagaag	615
Db	427	gccaaagccctctacactgaagaggtltgagcatltaaaagctccaaagaaagtcagaag	486
OY	616	cgtcttctccggagcctatgagctcatatgcttgggcttctatlyggttccaaacttgaagacgg	675
Db	487	cgtcttctccggagcctatgagctcatatgcttgggcttctatlyggttccaaacttgaagacgg	546
OY	676	ggcagcgggtgctgtatgc	693
Db	547	aacacagggggcgtatgc	564

RESULT	5
BE288040	
LOCUS	
DEFINITION	BE288040 1050 bp mRNA EST 13-JUL-2000
ACCESSION	60109472601 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE3489342 5' ,
VERSION	BE288040
KEYWORDS	BE288040.1 GI:9167462
SOURCE	EST.
ORGANISM	house mouse.
REFERENCE	Mus musculus
AUTHORS	Emmalyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL	1 (bases 1 to 1050)
COMMENT	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .
	National Institutes of Health, Mammalian Gene Collection (MGC)
	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Tel.: (301) 496-1550
	Email: Robert.Strausberg@nih.gov
	Tissue procurement: Lothar Hennighausen Ph.D., Robin Humphreys
	CDNA Library Preparation: Life Technologies, Inc.
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Incyte Genomics, Inc.
	Clone distribution: MGC clone distribution information can be
	found through the I.M.A.G.E. Consortium/LLNL at:
	http://image.llnl.gov
	Plate: U1AM8530 row: c column: 07
	High quality sequence stop: 661.
FEATURES	Location/Qualifiers

```

FEATURES
SOURCE
Location/Qualifiers
1. 1050
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3489342"
/clone_lib="NCI CGAP Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/ncbi="Organ: mammary; Vector: PCMV-SPOrt6; Site:1: Salt:
Site:2: Nci1: Cloned unidirectionally. Primer: Oligo dt.
library constructed by llfe Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

```

BASE COUNT	285 a	272 c	328 g	164 t	1 others
ORIGIN					
Query Match	22.3%	Score 502.4;	DB 34;	Length 1050;	
Best Local Similarity	87.0%;	Pred. No. 2.1e-100;			
Matches 600;	Conservative 0;	Mismatches 81;	Indels 9;	Gaps 4;	
QY	745	aacaacctgcgtattacaagcgcacccatcaagtaactcgtgtgtagctcgtgtagaactac	804		
Db	1	AACATCTGCGATTACACGCATCCTCAAGTCGTGGCGGACCTGGGCTTACAGACACTAC	60		
QY	805	caggagccccctgctgcgtcttctccttgtagaggagaccctgttcaacacaaactgcacc	864		
Db	61	CAGGCACCGCTGTGCGCTTCTTCTTGAGGAGACCCATGTAACACACAAATGCCACAC	120		
QY	865	gtgcgcacagagtgccctgagactctccctgttcctgtcgtctgcgcgcagcagccg	924		
Db	121	GTTCCCCAGAGTGGCCCTGGACTACTTCTGTTGCCGTGCGTGGCCGACACAGCCGG	180		
QY	925	gagctgtgtactcttgcctbgyggagactcaagcctcgcgcgagagtltgtctcgyggccc	984		
Db	181	GAGCTGTCGACATTGTGCTTGGGAGACACTCAAGCCTCGCCGAGAGTTGTGTGGGGCCC	240		
QY	985	cgtgacaagctgcgagagattcaagccccaagacataccccaagcacatgcagagaccag	1044		
Db	241	CGTGACAAGCTGCGGAGGTTCAAGGCCCCACAGATATATCCCGCCCACTGATGGACTAGGG	300		
QY	1045	caggcagataaagatgaggggtcccaaggagaccctccaaagaggtcgtgcacccagggtcg	1104		
Db	301	CAGGCAGATTAAGATGAGGGGCCCGGGGACCCCTCCCAAGAGGCTGGCACCAAGGGTGG	360		
QY	1105	acctgtgatatctgaaggagaccctgagtgtggagcagttggaacacagctgagagatccctca	1164		
Db	361	ACCTGTGATCTGGAAGGGACCTGAGTGGGACATGGAACAGCTGAGGATCTCTACAG	420		
QY	1165	ctgaacacaagaccctcagatgtggggaaacctgtgattggyaacagagagatgaagctaa	1224		
Db	421	TTGACCGCAAAAACCCACGAGATGTGGGAACCTTGATGGGAGCAAAAGCATGAGCTAAG	480		
QY	1225	tcacctgattcccaagagagacgaagaaaggaagtgtgggggggaaacagagagagcaggtc	1284		
Db	481	TCGCCGAGTCCCAAGAAAGCA - GAAGAGGAAGTGGAGGGGAACAGCAGGACGAGTCC	538		
QY	1285	cca -ggggagagcagatccccaaggtgtctctgaagtagaagaatgtcccttaacctga	1343		
Db	539	CCAGGGGGAGCCAGATCCTCAGGGGTGTCTCTGAGTAGAGAAAATGGCC - TCAACTTGA	596		
QY	1344	ggaggtgtcccttaagccctatcagcagagagagccagggagagctgaacccgtctcgt	1403		
Db	597	GGGGGTGCCCTCCACACC --- TACAGCAGGAGGCCCTGTAGCTGACAAAGCCTGTCTTGT	652		
QY	1404	ggcagaggtgtgctaattgaggtagaagcgcg	1433		
Db	653	GGCTAGTGCTTAATGAGTAGTACGAACCG	682		
RESULT	5				
LOCUS	AW762644	500 bp	EST	04-MAY-2000	
ACCESSION	U45402.y1	NCI-CGAP	Man3 mus musculus cDNA clone IMAGE:3155091	5'	
VERSION	AW62644		similar to TR:096029 096029 7-60. ;	mRNA sequence.	
KEYWORDS	AW762644.1	GI:7694574			
SOURCE	EST.				
ORGANISM	house mouse.				
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 500)				
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).				
	Tumor Gene Index				

JOURNAL
COMMENT

Unpublished (1997)
Other ESTs: urf6402.x1
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
Image.llnl.gov/image/html/resources.shtml

MG1:1057847
Seq primer: -40RP from Gibco
High quality sequence stop: 438.
Location/Qualifiers

FEATURES

source

1..500
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3155091"
/clone_1lb="NCI-CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

BASE COUNT 139 a 125 c 148 g 88 t
ORIGIN

Query Match 20.9%; Score 470.8; DB 24; Length 500;
Best Local Similarity 96.6%; Pred. No. 1.7e-93;
Matches 481; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
QY 136 tccctcttcgagatgagacccgagctcgattccacctggagagagagcgag 195
DB 3 tccacgctctcagatgacacacccggagtcgattccacctggagagagcgag 62
QY 196 gagagatgagagatgagcagcgagatgatacagacgagtgagacgagcgagatgac 255
DB 63 GAGGACGGCGAGGAGATGAGCGAGCGGATGATACGACCGATGAGGACGCGGAGCATGAC 122
QY 256 ggcgcgcggagagagcgagcgagcgctgttcacgctcagagatgacagggtaaccgaac 315
DB 123 GCGGACGGGAGGAGGACGGCCAAAGCCTGTCTCAGTGTAGATGACAAAGGTACCGAAGC 182
QY 316 tggcgtgctatgacgagatgacaaagataccggcacaactaccggatttacagatcaa 375
DB 183 TGGCGTCTATGACAGGACATGCAAAAGATACCGGCGACACTACCGGATTACGATCAA 242
QY 376 gactgcaatggtgacatgtgcaactgagcttctacaaaatgagatctgcttcacgca 435
DB 243 GACTGCAACGGAGACATGTGCAACCTGAGCTTACAAAATGAGACTGCTTCCACGCA 302
QY 436 aatggagcttcacacgagacattcttcagacatgagaaagaaagaaacttgactcttgaa 495
DB 303 AATGGCTTCTCATTTAGAGGACATCTTCAGAACTGGAAGCACTATGACCTCTCGAA 362
QY 496 gagaatacctctacatcagtgctcttctctgcyggaacacagagtgaaactgycac 555
DB 363 GAGATACACTCTCACTCACTGCTGCTTCCCTGAGGGAACAGAGTGAAGTGGCAT 422
QY 556 gccaaagccctcaccctgaaagaggttgagggcatttaagcttcaaggaagtcagagag 615
DB 423 GCCAAGCCCTCACTCACTGAGAGGTTGAGGCAATTAAAGCTCCAAAGGAAGTCAGAGAG 482
QY 616 cgtcttgcgggacctat 633
```

DB 483 CGCTGTGTCGGGCGCTAT 500

RESULT 7

LOCUS AM221684 506 bp mRNA EST 25-JAN-2000
DEFINITION uc037c01.y1 NCI-CGAP_Mam6 Mus musculus cDNA clone IMAGE:2631648 5', similar to TR:096029 096029 7-60. ;, mRNA sequence.
ACCESSION AM221684
VERSION AM221684.1 GI:6751228
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbfp/image/image.html

MG1:1025156
Seq primer: -40RP from Gibco
High quality sequence stop: 423.
Location/Qualifiers

FEATURES

source

1..506
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:2631648"
/clone_1lb="NCI-CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 141 a 125 c 149 g 91 t
Query Match 20.6%; Score 462.8; DB 21; Length 506;
Best Local Similarity 95.6%; Pred. No. 9.6e-92;
Matches 476; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
QY 136 tccctcttcgagatgagacccgagctcgattccacctggagagagagcgag 195
DB 9 tccacgctctcagatgacacacccggagtcgattccacctggagagagcgag 68
QY 196 gagagatgagagatgagcagcgagatgatacagacgagtgagacgagcgagatgac 255
DB 69 GATGACGGGAGGAGATGAGCGAGCGGATATACGACCGATGAGGACGCGGAGCATGAC 128
QY 256 ggcgcgcggagagagcgagcgagcgctgttcacgctcagagatgacagggtaaccgaac 315
DB 129 GCGGACGGGAGGAGGACGGCCAAAGCCTGTCTCAGTGTAGATGACAAAGGTACCGAAGC 188
QY 316 tggcgtgctatgacgagatgacaaagataccggcacaactaccggatttacagatcaa 375
DB 189 TGGCGTCTATGACAGGACATGCAAAAGATACCGGCGACACTACCGGATTGACGATCAA 248
```

Oy	376	gactcgaatggagacatggcaaccaggcttcatacaaaatgatgtctccagcca	435
Db	249	GACTCGACAGGAGACATGTGCACAACCTGAGCTTCACAAAAAATGAATCTGCTCCAGCCA	308
Oy	436	aattgggtctcatcgaggacatcttccagaactggaagaacaactatgacctcttgaa	495
Db	309	AATGGCTTCTCATTTGAGSAGCAATTCTTGAGAAGCTGGAAAGAACACTATGACCTCTGGAA	368
Oy	456	ggagatcacctcaatcacagtgcgtgtttctctctgcggggaaccagagatgaactggc	555
Db	369	GAGAATCATTCTTAATCCAGTGGCTGTTCCTCCGAGGGAACGAGTAGTAACTGGCAT	428
Oy	556	gccaaagccccctaacctcgaaggaggttgagcatttaaaagctccaagaagttagag	615
Db	429	GCCAAAGCCCCCTACACTGAAGSAGGTGTATGTCATGTTAAAGTCCAGAGACTCAGAGAG	488
Oy	616	cgtctgtccggagccat	633
Db	489	CGTCTGTCTGGGCCCTAT	506
RESULT	8		
LOCUS	AM476433		
DEFINITION	ug75g05.v1 NCI-CGAP Lu33 Mus musculus CDNA clone IMAGE:2936984 5'	EST	24-FEB-2000
ACCESSION	AM476433		
VERSION	AM476433.1	GI:7046539	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 527) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
JOURNAL	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.lnl.gov/db/rp/image/image.html		
COMMENT			
FEATURES			
source	Seq primer: -40RP from Gibco High quality sequence stop: 483. Location/Qualifiers 1..527 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:2936984" /clone_id="NCI CGAP Lu33" /tissue_type="Pooled lung tumors" /lab_host="DH10B (phage-resistant)" /note="Organ: Lung; Vector: pRT33-Pac (Pharmacia) with a modified polylinker. Site.1: NotI; Site.2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - 0190(dT) primer [5]. TGTATCACATCTGAAGTGGAGCGCGCCGCTCTGTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT33 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	118 a 146 c 171 g 91 t 1 others		

Query Match	19.6%	Score 440.6;	DB 22;	Length 527;
Best Local Similarity	92.0%;	Pred. No. 7.5e-87;		
Matches 486;	Conservative	0;	Mismatches 40;	Indels 2;
Qy 754	cgatatacagcagatc-ctcaagtcactctgtgtgagctgtgtgttagaacaactaccagcacc	812		
Db 1	CGATTATACACGCAATCTCCAACTGCTCCCTGGGAGAGTGGGCTTAGAGCACTACACAGGAC	60		
Qy 813	ccgtgtccgcttcttcctcctcggagagacccctgttaagacaacacgtccaggtgtgcgca	872		
Db 61	GCTGTGCGCTTCTTCTCCGAGAGAACCTTATGACAGCAAACTGCCAGGCTTCGCCA	120		
Qy 873	gagtgcccttgactactctctgtctgtctgtcgtctgcgcacacagccgcgcgagacttgt	932		
Db 121	GAGTCCCTCGACACTCTCTTCTGCGGTGGCGTCCGCCGACACAGCCCGGAGACTGT	180		
Qy 933	gtactcttgccttgaggacacttcaagcctcgcgcagagattgtctgtgtgtgtcccgatgaca	992		
Db 181	GCACCTTCTCTTGGGAGCACTTCACACCTCCGCCGAGAGTTGTCT- GGGGCCCCCGTGACAA	239		
Qy 993	gctgtggagattcaagccccagaccatccaccagcacttgacagcagcagcagcagca	1052		
Db 240	GCTGGGAGGTTCAAGCCCCACAGCAATATCCCGGCACCTGAGGAGCTAGGGGACAGGACGA	299		
Qy 1053	taaaatagagagctccagagaccctcccaagagctgtgaccccgaggtctgagactgtgtg	1112		
Db 300	TAAAGATAGAGGCCCCCGGGGACCCCTCCCAAGAGGCTGGACACCGAGGCTGGACTGTGG	359		
Qy 1113	atcttgaaaggagcctagctgt	1172		
Db 360	ATCTGGAAGGAGGACCTGAGTGGGAGCAGTGAACACACTGAGATCTCATCTGTGTGAGCGC	419		
Qy 1173	aaagccctcagatgt	1232		
Db 420	AAACCCCAAGGAGTGTGGGAACTTGTGATGGGAGACCAAGGATGAGGCTAAGTCCCGAG	479		
Qy 1233	tcaccaagagagacagaagaagaagt	1280		
Db 480	TCCCAAGAGAGCAAGAGAGAGCAATTTGAGAGGGAACACNGCATGAGCA	527		
RESULT 9				
AI1786854	447 bp	mRNA	EST	02-JUL-1999
LOCUS	u132d09.y1	Sugano mouse kidney	klia	Mus musculus
DEFINITION	IMAGE:1921643 5', mRNA sequence.			
ACCESSION	AI1786854			
VERSION	AI1786854.1	GI:5334570		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
	1 (bases 1 to 447)			
	Marra,M., Hallier,L., Kueba,T., Martin,J., Beck,C., Wylie,T.,			
	Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Peterson			
	,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R.,			
	Waterston,R. and Wilson,R.			
	The WashU-NCI Mouse EST Project 1999			
	Unpublished (1999)			
	Other.EMBL: u132d09.x1			
	Contact: Marra M/WashU-NCI Mouse EST Project 1999			
	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: mouseest@wustl.edu			
TITLE	This clone is available royalty-free through LINT ; contact the			
JOURNAL	IMAGE Consortium (info@image.llnl.gov) for further information.			
COMMENT				

OY	680	cggggtgcgtatgccgtgtccacagaaattcccacgcgcgtctccaacaattgaaacagcacca	739
Db	570	CGGGACGCGTGGGCCGAGACAGAACTACCAAGACGCCTTCAGAACTGAACCTGGCCGA	629
OY	740	gccacaacaacctggtatllatacgcacatccctaagaactggtgtagtgcgttagaac	799
Db	630	GCCAAACAACACTCCTCGCATCACACGCANCTCAAGTCGTGGTGAGCTGGGCGTCGAGC	689
OY	800	actaccag--cacccttgctcgctttctctgtgaagagagacccttgcacagacaact	857
Db	690	ACTTCAGAGGCGCCGCTGGGTCTCGCTTTCTCTCGAGGAGAAGCGTGTGTCGGCGGTGA	749
OY	858	gcccaagctgt-----cgcaagatgccccgg--actacttcctgttcgtgtgcctgccc	910
Db	750	GCTGCCGGGGGGTGCCTGGCAGAGATGCTTCGAGACTTACTTAATGATGCGCGCTGCCTGCC	809
OY	911	ggcaccagcgcgcggaggactgtgtacttcgtccgtggagagcaactcaagctcg	962
Db	810	GACACCAGGCGCTGCAGGCTGTGGCCCTTGCGCTTGGAGACCTTTGGGGCCCC	861
RESULT	11		
AM538283/c			
LOCUS			
DEFINITION	AM538283	508 bp	mRNA EST 02-AUG-2000
ACCESSION	C0105F08-3	Mouse E7.5 Extraembryonic Portion cDNA Library Mus musculus CDNA clone C0105F08 3', mRNA sequence.	
VERSION	AM538283		
KEYWORDS	AM538283.1	GI:7180700	
SOURCE	EST.		
ORGANISM	house mouse.		
	Mus musculus		
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
COMMENT	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 508)		
	Tanaka,T.S., Jaradade,S.A., Lim,M.K., Kargui,G.J., Wang,X., Grabovacek,I.II, Becker,K.G. and Ko,M.S.H		
	Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)		
	Contact: George J. Kargui		
	Laboratory of Genetics		
	National Institute on Aging/National Institutes of Health		
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA		
	Email: cdna@igsun.grc.nia.nih.gov		
	Plate: C0105 row: F column: 08		
	Seq primer: -21M13 Forward		
	High quality sequence stop: 508		
	POLYA-Yes.		
FEATURES			
SOURCE			
	Location/Qualifiers		
	1..508		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="C0105F08"		
	/clone_id="Mouse E7.5 Extraembryonic Portion CDNA library"		
	/sex="unknown"		
	/dev_stage="7.5dpc Embryo"		
	/lab_host="DH10B"		
	/note="Vector: pSPOR1 (Gibco/BRL Life Technology); Site_1: SalI; Site_2: NotI; Total RNAs were extracted from 5 EPC. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dt) primer [NotI primer-adaptor from GibcoBRL]		
	[5'-pGACTACTTCTAGATCGGAGCGCGCCCTTTTTTTTTTTT-3']		
	from 0.8ug of mRNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-Sal3 (include Sal1 sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range hignfidelity PCR using Takara's Ex Taq		

Query Match	17.4%	Score 391.8;	DB 22;	Length 508;
Best Local Similarity	90.38;	Pred. No. 4.1e-76;		
Matches 465; Conservative	0;	Mismatches 42;	Indels 8;	Gaps 4;
1689 agagagaccacaacagccaggtcgcggccagagagaccacaacagccaggtcgcggccagagga 1748				
Db 508 AGAGAGACCCCAAAAGCCAGGTGGAGCCAGAGACCACCAAAAGCCAGGTGGGGCCAGAGGA 449				
QY 1749 cccaacagccaggtgtyggggccagagcaagctgctctaaagccctgttggagagacc 1808				
Db 448 CCCCCAAAGCCA---GGTGGGGCCAGACAGACGTGCTCTTAAGAGCCCTGGGGAGAGATCC 392				
QY 1809 tgaccttcaacctatgynaacctcaagtgaatgaatcagagagagatgttgcaagatattgagc 1868				
Db 391 TGACCTGACACTACAGGAAACCTCAATAGAGAGCAAGAGAGTGGCGAAGATTGAAGGC 332				
QY 1869 ctctgtcgaaccccccaagccttaagaggtgatcctcagctctactcaagccccactgcaagg 1928				
Db 331 CTCGTGTAACCCCAAGACCTTAGAGGTACCCTCACTCTAATTAAGCCACTGCAAGG 272				
QY 1929 ggttctcgtagtcacagagctcgcgtgaagctctctcttcttggtyggcccaagctgtgacctct 1988				
Db 271 GGTTCCTG-TCCAGAGAGCTGTGTGTTGGCTCTCTTGTGGTCCCAAGTGTGCTGCTCT 213				
QY 1989 cccatagtgctcatcagtgaggtggccacacagagagactgagagccctgctcctcagaagagcca 2048				
Db 212 CCTAGTGTGTCATGAAATGGCCACACAGAGGAGTGAAGGCCCTGCCCTCAGAGGAAGGCCA 153				
QY 2049 aggccttcaagaacctctctcaactcagctgctcctcctcaactgcccctctgaagccctggc 2108				
Db 152 AGGCTTGTAGAGGCTCTCTTACCTCACTGAG---TCTCCACAGTGCCTCTGTAGGCCCTGTG 96				
QY 2109 ttgtgatcaagacccttaagaggtcctagagagagagggcctc-ttcattagctgtgtgcagaat 2167				
Db 95 TTGTGTAAGAGACCCCAAGAGGCTTAAGGGAGTGGGGCTTTTCTTCTGCTCATGCAAGT 36				
QY 2168 gaggcctttcttgataaactctcttaagctttgtc 2202				
Db 35 GAGGCCCTTCTGTGAATAAACCTTTAGGCTTTGTGTC 1				

polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gldco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang."

BASE COUNT	109 a	144 c	146 g	109 t
ORIGIN				

RESULT 12
A1578564/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

403 bp mRNA EST 05-APR-1999
UI-R-AAO-w1-g-01-0-UI-s1 UI-R-AAO Rattus norvegicus CDNA clone
UI-R-AAO-w1-g-01-0-UI 3', mRNA sequence.
A1578564
A1578564.1 GI:4562940
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 403)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB


```
|||||
Db 422 GAATGCGGAGCCACAACACCTCCGATCAGACGATCTCAAGTCGTGGTGA 481
QY 789 ggagcttagaacacacacagcagccctggtcgtcttc 839
Db 482 GGGCTCGAGCATTCAGCGCGCGTGTGCTTCCTTC 522

RESULT 14
LOCUS AM251730 386 bp mRNA EST 12-DEC-1999
DEFINITION UI-R-BJ0-aef-c-04-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone
ACCESSION AM251730
VERSION AM251730.1 GI:6595321
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Rattus.
REFERENCE 1 (bases 1 to 386)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized ventricle at 13 dpc library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-Yes.

FEATURES
source location/Qualifiers
1..386
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0-aef-c-04-0-UI"
/clone_1lb="UI-R-BJ0"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
library is a subtracted library derived from the UI-R-A01,
UI-R-A01, UI-R-A01, UI-R-A01, UI-R-A01, UI-R-A01, and
UI-R-A01 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the NotI site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB=UI-R-BJ0
TAG_TISSUE=ventricle at 13 dpc
TAG_SEQ=CAGCGA"

BASE COUNT 90 a 100 c 107 g 88 t 1 others
ORIGIN
Query Match 16.8%; Score 377; DB 21; Length 386;
```

```
Best Local Similarity 98.4%; Pred. No. 7e-73;
Matches 380; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1831 tcaagtgaatagtcagagagagtggtgcaagagatgagagctctctgaaccccaagact 1890
Db 386 TCACTGTGATAGTCAGAGGAGTGGCAAGAGATTGAGCCCTCTCTAAACCCCAAGCC 327
QY 1891 tagagtgacatcagtcagtcacacacacacacacacacacacacacacacacacac 1950
Db 326 TAGAGTGATCTCAGTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 267
QY 1951 ccgttagctctctctgtgtgcccacagtgctggtcctctccctagtggtcactgagtg 2010
Db 266 CCGTAGGCTCTTCTGTGTGTCGCCACAGATGNTGGCTCTCCCTAGTGTACTGAGTG 207
QY 2011 caccagagagagagagagcctgcccctcaggaagagagagagagagagagagagag 2070
Db 206 CACCAGAGGAGACTGAGGAGCCCTGCGCTCAGGAGGAGGAGGAGGAGGAGGAGGAG 147
QY 2071 ctcactgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2130
Db 146 CTCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 87
QY 2131 tagagagagagagagagagagagagagagagagagagagagagagagagagagag 2190
Db 86 TAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 27
QY 2191 ttagacttgcaaaaaaaaaaaaaa 2216
Db 26 TTAGACTTTTAAAAA 1

RESULT 15
LOCUS AM538292 529 bp mRNA EST 02-AUG-2000
DEFINITION C0105G09-3 Mouse E7.5 Extraembryonic Portion cDNA Library Mus
ACCESSION AM538292
VERSION AM538292.1 GI:7180709
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 529)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,
M.J., Pantano,S., Sano,Y., Piao,Y., Nagatsuma,R., Dol,H., Wood,W.H.
TITLE National Institute on Aging/National Institutes of Health
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnelson@nig.nih.gov
Plate: C0105 row: 6 column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 529
POLYA-Yes.

FEATURES
source location/Qualifiers
1..529
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C0105G09"
/clone_1lb="Mouse E7.5 Extraembryonic Portion cDNA
Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DHI0B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
```

5 EPC. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer [NotI primer-adaptor from GibcoBRL] [5'-PGACTAGTCTAGATCGGAGCGGCCCTTTTCTTTT-3'] from 0.8ug of mRNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang."

BASE COUNT 121 a 149 c 141 g 118 t
ORIGIN

Query Match 16.6%; Score 374.2; DB 22; Length 529;
Best Local Similarity 87.5%; Pred. No. 3.1e-72;
Matches 468; Conservative 0; Mismatches 58; Indels 9; Gaps 5;

QY 1672 aacagccaggtctgggc-cagagagaccacagccaggtcggccagagagaccacag 1730
DB 529 AAAAGCCAGGTGAGTCCAGAGACCCCAAGAGCTAGTGATCCAGATGACCCCAAG 470
QY 1731 ccaagtcggccagagagaccacagccaggtggtgggccaagagcaagctgctctaa 1790
DB 469 CCAGGTGGGGTCAGAGGACCCCAAGCTA---GGTGGGCTAGAGTAAGCTGCTTAA 413
QY 1791 gagccctgtgagagaccctgactctgacactatgggaacctgaatgagtcagagaa 1850
DB 412 GAGCATGGGGGAGATCCTGATTCACAGTAAGAACTCAATGAGTACAGAGGA 353
QY 1851 gttgcaaggtattgagcctcgtcgaaccccaagccttagagtgatctcagtcct 1910
DB 352 GTTGCAGAGGATTGAGGCTCTGTTGAACCCCAAGCCTTAGAGTAACCTTCAGTCT 293
QY 1911 actcagccactgacaggggttctgtagtcagagctcgcgtaggctctctgtgac 1970
DB 292 AATTAGCCCACTGACAGGGGGTTCTG-TCCAGAGCTCTGTTGTGCTCTTGTGTC 234
QY 1971 cccaagatgctgagcctctcctagtgatcagtgagtgagccagagagactgagccc 2030
DB 233 CCCACAGTGTGGCTCTCCCTAGTGTGTCAGTGAAGTGCCACAGAGGAGTGAAGGCC 174
QY 2031 tgcctcagggagggcagagcctcagaaacctcctaacctcactgtctcctcagc 2090
DB 173 TGCCCTCAGGGAGAGGCCCAAGGCTTCAGAGGCTCTTACCTCACTGTG--TCTCCAC 117
QY 2091 tgcctcagggcctcgtgtgacacagccctaaaggtctagagagagggcctc-ttc 2149
DB 116 TGCCCTCAGAGGCTGTGTGTGAACAGCCCTAAGGGCTAAGGGGATGGGCCCTTTT 57
QY 2150 attagtcggtgagcagtgagccttctcgaataaactcttagactttgcaa 2204
DB 56 CTTCGTGTGATGCCAAGTGAGGCTTTTCTGAAATAACTTTAGCTTTGTGCTA 2

Search completed: October 14, 2000, 03:41:44
Job time: 7587 sec

US-07-884-811-15
; Sequence 15 Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-884-811-15

Query Match 3.6%; Score 57.2; DB 1; Length 10596;
Best Local Similarity 43.2%; Pred. No. 0.0013;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

QY 718 caggcgccactgtccgcttctctctgagagagcgtgtgagggagctgtccggg 777
DB 2271 CAGGACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2330
QY 778 gtgcgacagagtcctctgactactcaagttcgcgtgtgcgcgcgaacacagcgccg 837
DB 2331 GGGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2390
QY 838 cagctgtgtcacttcctctgtgagacacttcgcgcgcgcgcgcgcgcgcgcgcgcgc 897
DB 2391 GAGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2450
QY 898 caagaaacgtcgtgaggttcaagccagctctctgcgcgcgcgcgcgcgcgcgcgcgcgc 957
DB 2451 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2510
QY 958 aaggtgag 1017
DB 2511 GGGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2570
QY 1018 accctgtggccagacatagcaaggtgtggtgaggtgagagagagagagagagagagag 1077
DB 2571 GAGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2630
QY 1078 agcgtgag 1137

DB 2631 GGGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2690
QY 1138 cagcgagagatagcccgagccttaagcccccaagagagcaagaagagagagagagag 1197
DB 2691 GAGGCGAGAGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2750
QY 1198 ctgagccgagcgagagcagccgcacagagccctcagatgtcctcagaggtgag 1257
DB 2751 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2810
QY 1258 aagatcgctcttaatttgagaggtgtgctcctcagcagggcagccttaagagcgagc 1317
DB 2811 GGGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2870
QY 1318 cagaagtggtgcgtcaggaacctgtggag 1347
DB 2871 CAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2900

RESULT 11

US-07-885-971-15
; Sequence 15, Application US/07885971
; Patent No. 5328837

; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/885,971

FILING DATE: 19920518

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: 779

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-3216

TELEFAX: 415/952-9881

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 10596 bases

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

US-07-885-971-15

Query Match 3.6%; Score 57.2; DB 1; Length 10596;
Best Local Similarity 43.2%; Pred. No. 0.0013;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

QY 718 caggcgccactgtccgcttctctctgagagagcgtgtgagggagctgtccggg 777
DB 2271 CAGGACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2330
QY 778 gtgcgacagagtcctctgactactcaagttcgcgtgtgcgcgcgaacacagcgccg 837

```

Db      2331 GGGCAGGAGCGAGGAGGGGCGAGGACAGAGAGGAGGGCCAGAGGGGCAGAGAGGGCGAG   2393
Oy      838 cagctgtgtgcaattcgcctcttggaagacacttcgggcccgctgcaagtctgttgyggcc   897
Db      2391 GAGCAGGAGGGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAGGGGCCAGAGGACAGAGAGGGG   2450
Oy      898 caagaacaagctgcygaggttccaagcccagctctctctgcgcatactgcgtctgaggtctccagg   957
Db      2451 CAGGAGGGGCGAGGAGGGGCGAGGAGCGAGGAGGAGGAGGGGCCAGAGGACAGAGAGAGGGGCCAGAG   2510
Oy      958 aagttgagtgagagaaagcccccgaggaccacagagcgagccagacccagaggttcg   1017
Db      2511 GGGCAGGAGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCCAGAGGACAGAGAGAGGGCGAG   2570
Oy      1018 acctgtgycaccaagacataagcaaaggttggyggcaggytgcacaggyggccccagaccag   1077
Db      2571 GAGCAGAGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAG   2630
Oy      1078 agctgtgagcccccagagatctgcyggaacccctcgyagagyaagcaccaggggatgtgagcaggggc   1137
Db      2631 GGGCAGAGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG   2690
Oy      1138 cacgcygaagatatagycggcgagcccttaagccccaaagagagcaagaagaggaagcttgtg   1197
Db      2691 GAGGCGGAGAGCGAGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG   2750
Oy      1198 ctgagccgcygcygagcacgcccaccacagagccagagcgcctcagaagtgtcctcagaagtgtgag   1257
Db      2751 CAGAGAGGGCGAGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAGGGGCCAGAGAGAGGAGAG   2810
Oy      1258 aaagtgcctctgaaattgtagaggygtgtgcctcagcagagcgagcctaagagcgagaac   1317
Db      2811 GGGCAGAGAGGGGCGAGGAGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAG   2870
Oy      1318 caggaaagtgcgcytcacagagacccttgtyggag   1347
Db      2871 CAGGAGCAGAGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAG   2900

RESULT 12
US-087-783A-15
Sequence 15 Application US/08087783A
Patent No. 554/856
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
TITLE OF INVENTION: HEPATOCTYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 Inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/087,783A
FILING DATE: 13-Jul-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-May-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885971
FILING DATE: 18-May-92
ATTORNEY/AGENT INFORMATION:
NAME: Marschang, Diane L.
REGISTRATION NUMBER: 35,600
REFERENCE/DOCKET NUMBER: P0755779P1

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5416
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-087-783A-15

Query Match      3.6%; Score 57.2; DB 1; Length 10596;
Best Local Similarity 43.2%; Pred. No. 0.0013;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

Oy 718 caggcgccacgtgtctccgttcttctcttgagagaaagctgtgtgctgagctgtccggg 777
    |||
Db 2271 CAGGAGCCAGAGAGAGAGGGGCAAGAGAGAGAGAGGGGCGAGAGGGGCGAGAG 2330
    |||

Oy 778 gtgcgcagagatgtccctgacttcaattgttctgcgctgtgcctgcacacagcgccg 837
    |||
Db 2331 GGGCGAGAGCAGAGAGAGAGGGGCAAGAGAGAGAGAGGGGCGAGAGGGGCGAG 2390
    |||

Oy 838 caagctgtgtcaccttcgctcttgagagacttccgcccgcctgcaagtctgtctggggccc 897
    |||
Db 2391 GAGCAGAGAGAGAGGGGCAAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGG 2450
    |||

Oy 898 caagaacaagctgtcggaggttcaagccacgtctctgtccgcatcgtctgagaggtctcaag 957
    |||
Db 2451 CAGGAGGGGCGAGAGAGGGGCAAGAGAGAGAGAGGGGCGAGAGGGGCGAGAG 2510
    |||

Oy 958 aaggtgtgagagagagagagagagagagagagagagagagagagagagagagagagagag 1017
    |||
Db 2511 GGGCAGGAGCAGAGAGAGAGGGGCAAGAGGGGCGAGAGGGGCGAGAGAGAGAG 2570
    |||

Oy 1018 acctgtggccagagacatagcaaggtgtggggcagaggtgagcagaggggcccacagcag 1077
    |||
Db 2571 GAGCAGAGAGGGGCAAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAG 2630
    |||

Oy 1078 agcgtgtgagagagagagagagagagagagagagagagagagagagagagagagagag 1137
    |||
Db 2631 GGGCGAGAGGGGCGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGAG 2690
    |||

Oy 1138 caccgggagagataggtccggagcctctaagcccccaagagagcaagaagaagatgtgag 1197
    |||
Db 2691 GAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGGGGCGAGAGAG 2750
    |||

Oy 1198 ctgagccggcggagacagcccccacagacagagccctcagagtgtcctcagaggttgag 1257
    |||
Db 2751 CAGGAGGGGCGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAG 2810
    |||

Oy 1258 aagatcgtctgaatttggaggggtgtgtccctcagccagagcagagctcagagacggggacc 1317
    |||
Db 2811 GGGCGAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGGGGCGAGAG 2870
    |||

Oy 1318 cagggaagtggtgcgttcaggacccttgaggag 1347
    |||
Db 2871 CAGGAGCAGAGAGGGGCGAGAGGGGCGAGAG 2900
    |||

RESULT 13
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
;

```

CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194.088B
FILING DATE: 09-FEB-1994
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Gallegos, R. Thomas
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: 755D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-2614
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-088B-15

Query Match 3.6%; Score 57.2; DB 1; Length 10596;
Best Local Similarity 43.2%; Pred. No. 0.0013;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

OY 718 caggcgccactgttcgcttcttcttgagagacgtgtgctgagagctgctgag 777
DB 2271 CAGGAGCAG 2330
OY 778 gtgcgagagagtgctgctgacttcttctgctgctgctgctgctgctgctg 837
DB 2331 GGGCAG 2390
OY 838 cagctgtgagcttctgctgagagcttctgctgctgctgctgctgctgctg 897
DB 2391 GAGCAG 2450
OY 898 caagacaagctgcgaggttcaagcttctgctgctgctgctgctgctgctg 957
DB 2451 CAGGAG 2510
OY 958 aaggtgag 1017
DB 2511 GGGCAG 2570
OY 1018 accctgtgag 1077
DB 2571 GAGCAG 2630
OY 1078 agctgtgag 1137
DB 2631 GGGCAG 2690
OY 1138 cagcgag 1197
DB 2691 GAGGAG 2750
OY 1198 ctgagcgag 1257
DB 2751 CAGGAG 2810

OY 1258 aagatcgtctgaatttggagagagagagagagagagagagagagagagagag 1317
DB 2811 GGGCAG 2870
OY 1318 caggag 1347
DB 2871 CAGGAG 2900

RESULT 14
US-08-194-087-15
Sequence 15, Application US/08194087
Patent No. 5879910
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194.087
FILING DATE: 18-MAY-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 779
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-087-15

Query Match 3.6%; Score 57.2; DB 3; Length 10596;
Best Local Similarity 43.2%; Pred. No. 0.0013;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

OY 718 caggcgccactgttcgcttcttcttgagagacgtgtgctgagagctgctgag 777
DB 2271 CAGGAGCAG 2330
OY 778 gtgcgagagagtgctgctgacttcttctgctgctgctgctgctgctgctg 837
DB 2331 GGGCAG 2390
OY 838 cagctgtgagcttctgctgagagcttctgctgctgctgctgctgctgctg 897
DB 2391 GAGCAG 2450
OY 898 caagacaagctgcgaggttcaagcttctgctgctgctgctgctgctgctg 957
DB 2451 CAGGAG 2510

Mon Oct 16 10:08:49 2000

us-09-431-843b-7.rni

Page 10


```

: REFERENCE/DOCKET NUMBER: 30472/114 IMM0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match
Best Local Similarity 3.3%; Score 73.8; DB 1; Length 7218;
Matches 15; Conservative 235; Mismatches 137; Indels 0; Gaps 0;

QY 1408 aggtgtgctaatgtggtgaagaagcgaggaagtgaggaagggtcgaggtgatgga 1467
DB 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 1468 gtagcagtaacactcaaatgacagccagtcgctcctaccctcagagtgatcct 1527
DB 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 1528 gaggccaaagaatgaggaatgagccagagactcaaaacagccaggtgagcagaagat 1587
DB 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 1588 tccaaaagccagtggtgagccgagagatccaacagccggtgaggtctgagagccaac 1647
DB 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 1648 agccagtcggtgagcaggaagccaaacagccagtcggtgagcaggaagccag 1707
DB 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 1708 gtcgagcaggaagccaaacagccagtcggtgagcaggaagccaaacagagtggtg 1767
DB 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 1768 gggcagagcagctgctcctaagac 1794
DB 1071 RRRRRATCGAAGCTCCTCGACTGC 1045

RESULT 5
US-08-993-228-9
: Sequence 9, Application US/08993228
: Patent No. 5976838
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallie, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Weirberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Agostino, Michael J.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESSES:
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/993,228
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sprunger, Suzanne A.
: REGISTRATION NUMBER: 41,323
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8284
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4582 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-993-228-9

Query Match
Best Local Similarity 3.1%; Score 69.6; DB 4; Length 4582;
Matches 93; Conservative 1; Mismatches 20; Indels 2; Gaps 1;

QY 1 tgggtcagccagccagaggtgccccagtgaggactgtctcat--tctgacagct 58
DB 4371 TGGATATGACAACTCCCAAGGTCCAGAGTGAGTACTAGTATTACAGAGTGGCAGCT 4430
QY 59 gcacacatctgtcagtgaggaatgtcaggtctcactctcctcctcactc 114
DB 4431 GCACATACTGTTGTGTGAGGGAATATCCATTCCTCCTACTCTCCTCATATC 4486

RESULT 6
US-09-130-114-2/c
: Sequence 2, Application US/09130114
: Patent No. 5976807
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert A.
: APPLICANT: Robbins, Bassem B.
: APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
: FILE REFERENCE: 0867/1D903US1
: CURRENT APPLICATION NUMBER: US/09/130,114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 1931
: TYPE: DNA
: ORGANISM: EBNA
US-09-130-114-2

Query Match
Best Local Similarity 3.0%; Score 68; DB 4; Length 1931;
Matches 264; Conservative 0; Mismatches 305; Indels 3; Gaps 1;

QY 1197 ggaatggaaacagaggaatgaagctaaagtccttgatcccaaggagagaagaagaa 1256
DB 836 GGACGGGAGAGACGGGAGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 777
QY 1257 gttggaagggaacagaggaagcaggtcccaagggaagcagatcccaaggtgtcttga 1316
DB 776 GGACGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
QY 1317 gtagaagaatgtcccttaaccttgagaggtgtgccccttagccctcagcagagacc 1376
```

Db	719	CGGGAGGACACGAGGACCGGGAGGACGAGACGCGGAGACGCGGGAGACGGGAGAGAGA	660
QY	1377	caggagagctctgaaaccgcccctgtccctctgtgcacaggtgtgctaaatgagtaagaagcgyag	1435
Db	659	CGAGACACCGGGAGGACGAGACGCGGGGAGGACCGGGGAGACGCGGAGACGAGACGCGGGA	600
QY	1457	gaaaggttgaggaagaaggggtctgaggggtgaatgagtaagcaaaacaaatgcagcgag	1495
Db	559	GGAGGACGACGAGACCGGGAGGACCGGGGAGACCGGGGAGAGAGACGACGCGGAGACCGG	540
QY	1497	tgccctgtccctctaacccttcacaggtgtctcttgtaggcaccaaaagatctggtaatctggtcaga	1556
Db	539	GGAGGAGGACGAGGACCGGGGAGGAGGACGAGACGCGGAGAGACGCGGAGAGACGGGAGAGA	480
QY	1557	ggactcaaaacagccaggtctgggacagagatctcaaaagccaggtgtggccgagagatcc	1618
Db	479	GGACGAGGACCGGGAGGACCGGGGAGGAGGACGAGACGCGGAGAGACGAGACGGGGA	420
QY	1617	aaacaggtccaggtctgggtctgtaggggtcccaaaacaaacaaacaggtctgggtccaggtagaccacaaacag	1676
Db	419	GGACCGGGAGGACCGGGAGGAGGACGAGACCGGGGAGGAGGACGAGACGGGAGGACGG	360
QY	1677	ccaggtctgggtccagtaggacccaacaaacagccaggtctgggtccaggtagaccacaaacagccaggt	1736
Db	359	GGAGGACGGGAGGACCGGGAGGAGGACGAGACGCGGAGAGAGACGAGACGAGACGGGAGAGA	300
QY	1737	cggtgccagtaggacccaacaaacagccaggtgtgtgtg	1768
Db	299	CGAGGACGGGAGGAGGACGAGACGAGACGAGACGGG	268

```

RESULT 7
US-09-050-863-2
Sequence 2, Application US/09050863
Patent No. 6114111
GENERAL INFORMATION:
APPLICANT: Lao, Yang
APPLICANT: Hiang, Betty
APPLICANT: Payan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehr, Hobbach, Test, Abritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
City: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA

```

Query Match	2.7%	Score 60.4	DB 5	Length 2580
Best Local Similarity 51.9%				
Matches 136	Conservative 0	Mismatches 126	Indels 0	Gaps 0
QY 1519	gagtgctctgagcccaaaagatggaattggccagaggaactcaaacagccaggttgg			
Db 739	GAGGGGCAAGAGGGGACAGGACGAGGAGGGGGACAGAGCAAGAGAGGGGCGAGGAGG			
QY 1579	gagaaagatctcaaaagccaggttgggcccggaggatccaagaagccaggttgggttgg			
Db 799	CAGGAGGGGCAAGAGACAGAGAGAGGGGCGAGGACGAGAGGGGCAAGAGGGGCAAGAG			
QY 1639	gacccaacagccaggtctgggccaagagaccacaacacagccaggtctgggccaagagacc			
Db 859	CAGGAGGAGGGGCAAGAGGGGCGAGGAGGGGCGAGGACGAGGAGGAGGGGCGAGGAGG			
QY 1699	aacacagcagtgctgggccaagagaccacaacacagccaggtctgggccaagagaccacaagc			
Db 919	GAGGGGCAAGAGGGGCAAGAGACGAGAGGAGGGGCGAGGAGGGGCAAGAGGGGCAAGAG			
QY 1759	caggtcgttggggccaagccaag 1780			
Db 979	GAGGAGGGGCAAGAGCAGGAGG 1000			

```

RESULT      8
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: DamaJ, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903051
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: YEBNA
US-09-130-114-1

```

Query Match	2.7%	Score 60.4	DB 4	Length 5452
Best Local Similarity	51.9%	Pred. No. 4e-05		
Matches 136	Conservative 0	Mismatches 126	Indels 0	Gaps 0
QY 1519	gagttccctgagccccaagaagatgtagtattggtccagagagactcaaacacgaattctggg			1578
Db 2066	GAGGGGCGAGGAGGGGCAAGAGCAGAGAGAGGGGCGAGAGCGAGAGGAGGGGCAAGAGGGG			2007
QY 1579	gcagagagattcccaaaagcccaagtgtgggcccagagatcccaaacagcccaagtgtgggcttggg			1638
Db 2006	CAGGAGGGGCGAGGACGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG			1947
QY 1639	gaccnaaacacgcccgaattctgggtccagagagaccacaacagcccaagctctgggtccagagagacc			1698
Db 1946	CAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG			1887
QY 1699	aacagccagagttccgggtccagagagaccacaacagcccaagtgtgggtccagagagaccacaagc			1758
Db 1886	GAGGGGCGAGGAGGGGCGAGGAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG			1827
QY 1759	caggtgtgtgtgggtccagagcaag			1780
Db 1826	GAGGAGGGGCGAGGAGGAGGAGG			1805

```

RESULT 9
US-07-884-811-15
: Sequence 15, Application US/07884811
: Patent No. 5316921
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lohker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/884,811
: FILING DATE: 19920518
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dieger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 755.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-07-884-811-15

Query Match 2.7% Score 60.4; DB 1; Length 10596;
Best Local Similarity 51.9%; Pred. No. 5.3e-05;
Matches 136; Conservative 0; Mismatches 126; Indels 0; Gaps 0

QY 1519 gagtctctgaagcccaaaagtatggaattgagccagaggaactcaaacagccaggtttgg 1578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2319 GAGGGGCGAGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAGGAGGAGGAGGGG 2378

QY 1519 gcaaggaattccaaaagccaggttgagccgagagatccaacagccaggttggtgag 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2319 CAGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAG 2438

QY 1639 gaccacaacagccaggttcgagcagagagaccccaaacagccaggttcgagcagagaccca 1698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2439 CAGGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGGCGAGGAG 2498

QY 1699 aacagccaggttcgagcagagagacccaacacagccaggttcgagcagagacccaacagc 1758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2499 GAGGGGCGAGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAGGAGGGGCGAGGAG 2558

QY 1759 caggtgtgtggtggtccagagcag 1780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2559 GAGGAGGGGCGAGGAGCGAGGAGG 2580

RESULT 10
US-07-885-971-15
: Sequence 15, Application US/07885971

```

```

; Patent No. 5328837
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.,
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-885-971-15

Query Match      2.7%; Score 60.4; DB 1; Length 10596;
Best Local Similarity 51.9%; Pred. No. 5.3e-05;
Matches 136; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1519 gagtgccttgaggcccaaatgtggaatggcagagaagactcaaacagccagtgtgg 1578
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2319 GAGGGCGACGAGAGGGGCAGAGCAGAGAGAGAGGGGCCAGAGCAGAGAGAGAGGGG 2378

QY 1579 gcagaggaattccaataagccagagtggcgccgaagatccaacagccaggtgggtgag 1638
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2379 CAGGAGGGGCGACGAGCACGAGAGAGAGGGGCGACGAGCAGAGAGAGGGCCAGAGGAG 2438

QY 1639 gaccacaacacagccaggttcgggccagaggaaccaaacagccaggttcgggccagagacca 1698
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2439 CAGGAGGAGGGGCGACGAGAGGGGCGACGAGGCGCAGAGCAGAGAGGAGGCGACGAGAG 2498

QY 1699 aacacagcaggttcgggccagaggaaccaaacagccaggttcgggccagagaccaacagc 1758
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2499 GAGGGGCGACGAGAGGGGCAGAGCAGAGAGAGAGGGGCGACGAGAGGGGCGACGAGAGCAG 2558

QY 1759 caagtgtgtggggccagagcaag 1780
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2559 GAGGAGGGGCGACGAGCAGCAGAGG 2580

RESULT 11
US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.,
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
```

```

: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WinPatIn (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/087,783A
: FILING DATE: 13-Jul-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/884811
: FILING DATE: 18-MAY-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/885971
: FILING DATE: 18-MAY-92
: ATTORNEY/AGENT INFORMATION:
: NAME: Marschang, Diane L.
: REGISTRATION NUMBER: 35,600
: REFERENCE/DOCKET NUMBER: P0755779P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-5416
: TELEFAX: 415/952-9881
: TELE: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
:
: US-08-087-783A-15

```

```

Query Match      2.7%; Score 60.4; DB 1; Length 10596;
Best Local Similarity 51.9%; Pred. No. 5.3e-05;
Matches 136; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1519 gaagtctctgagcccaaaagatggaatgagccagaggaactcaaacagccaggtctggg 1578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2319 GAGGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2378

QY 1579 gcaagagattccaaaagccaggtctgggcccgaagatccaaacagccaggtctggg 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2379 CAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2438

QY 1639 gaccacaacagccaggtctgggcccgaaggaaccacaacagccaggtctggg 1698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2439 CAGGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2498

QY 1699 aacagccaggtctgggcccgaaggaaccacaacagccaggtctgggcccgaagga 1758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2499 GAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2558

QY 1759 caagtggtgggcccaggaag 1780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2559 GAGGAGGCGAGGAGGAGGAGG 2580

RESULT 12
US-08-194-088B-15
: Sequence 15, Application US/08194088B
: Patent No. 5580963
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,088B
: FILING DATE: 09-FEB-1994
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/884811
: FILING DATE: 18-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Gallegos, R. Thomas
: REGISTRATION NUMBER: 32,692
: REFERENCE/DOCKET NUMBER: 755D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-2614
: TELEFAX: 415/952-9881
: TELE: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-194-088B-15

```

```

Query Match      2.7%; Score 60.4; DB 1; Length 10596;
Best Local Similarity 51.9%; Pred. No. 5.3e-05;
Matches 136; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1519 gaagtctctgagcccaaaagatggaatgagccagaggaactcaaacagccaggtctggg 1578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2319 GAGGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2378

QY 1579 gcaagagattccaaaagccaggtctgggcccgaagatccaaacagccaggtctggg 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2379 CAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2438

QY 1639 gaccacaacagccaggtctgggcccgaaggaaccacaacagccaggtctgggcccga 1698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2439 CAGGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2498

QY 1699 aacagccaggtctgggcccgaaggaaccacaacagccaggtctgggcccgaagga 1758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2499 GAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2558

QY 1759 caagtggtgggcccaggaag 1780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2559 GAGGAGGCGAGGAGGAGGAGG 2580

RESULT 13
US-08-194-087-15
: Sequence 15, Application US/08194087
: Patent No. 5879910
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco

```



```

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPETER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,087
FILING DATE: 18-MAY-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dregler, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 779
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-087-15

```

	Query Match	2.7%	Score 60.4;	Dn 3	Length 10596;
	Best Local Similarity	51.9%	Pred. No.5,3e-05;		
	Matches 136;	Conservative	0;	Mismatches 126;	Indels 0; Gaps
QY	1519	gagtgctcctgagggcccaaaagtgtggaattggtgccagaggactcaaacagccagtttggg	1578		
Db	2319	GAGGGGCGAGGGGGCCACGAGCAGGAGCGGGGGCGAGGACAGAGAGGGCGAGGGGG	2378		
QY	1579	gcagaagatctccaagaacccaggttgggtgcgcgaagatactccaacagccaggttgggtcgag	1638		
Db	2379	CAGGAGGGGGCAGGAGCACGAGGAGGGGGCGAGGACGAGGAGGAGGGCGAGGGCGAGGAG	2438		
QY	1639	gaccacaacacagccaggttcgggtccagaaggaaccaaacagccaggttcgggtccagaaggaacca	1698		
Db	2439	CAGGAGGGGGGCCAAGAGGGGCGAGGAGGGGCGAGGACGAGGAAGGGCGAGGACGAGGAG	2498		
QY	1699	aacagccagcttcgggtccagaaggaaccaaacagccaggttcgggtccagaaggaaccaaacagc	1758		
Db	2499	GAGGGGCAGGAGGGGCGAGGACGACGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGG	2558		
QY	1759	caggttggttgggtccagaaggaag	1780		
Db	2559	GAGGAGGGGCGAGGACGAGGAGG	2580		

```

1 RESULT 14
2 PCT-US93-04648-15
3 Sequence 15, Application PC/TUS9304648
4 GENERAL INFORMATION:
5 APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melanie F.
6 TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
7 NUMBER OF SEQUENCES: 21
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Genentech, Inc.
10 STREET: 460 Point San Bruno Blvd
11 CITY: South San Francisco
12 STATE: California
13 COUNTRY: USA
14 ZIP: 94080
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648
FILING DATE: 19930517
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885871
FILING DATE: 18-MAY-92
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755,779P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-9216
TELEFAX: 415/952-9681
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
CCT-US93-04648-15

```

Query Match	2.7%;	Score 60.4;	DB 6;	Length 10596;
Best Local Similarity	51.9%;	Pred. No. 5.3e-05;		
Matches 136;	Conservative	0;	Mismatches 126;	Indels 0;
				Gaps 0;
QY 1519	gagtgtcctgagggcccaaaagatctggaatctggtgcagagagctcaaacagccaggtctggt	1578		
Db 2319	GAGGGGCGAGGGGGCGAGGACAGAGGAGGGGGCGAGGAGCGAGGAGGGCGAGGAGGGG	2378		
QY 1579	gcagaagattccaaagaacccaggtctggtgcgcggaagattccaaacagccaggtctggtgag	1638		
Db 2379	CAGGAGGGGCGAGGAGCGAGGAGGAGGGGCGAGGAGCGAGGAGGGGGCGAGGAG	2438		
QY 1639	gaccccaacagcccaagtcctggtccagaggaacccaacagccaggtctggtccagaggaacca	1698		
Db 2439	CAGGAGGGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGCGAGGAGGGGCGAGGAGG	2498		
QY 1699	aacacgcaggtctcggtgcagagagaccacaacagccaggtctggtccagaggaacccaacagc	1758		
Db 2499	GAGGGGCGAGGGGGCGAGGACGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG	2558		
QY 1759	caggtgtgtggtgcagaggaag	1780		
Db 2559	GAGGAGGGGCGAGGAGCGAGG	2580		

RESULT 15
 US-09-010-928B-3
 ; Sequence 3, Application us/09010928B
 ; Patent No. 5984099
 GENERAL INFORMATION:
 APPLICANT: Lewis, Randolph Y
 APPLICANT: Hayashi, Cheryl Y
 TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
 TITLE OF INVENTION: CODING THEREFOR
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
 STREET: 8110 GATEHOUSE RD, SUITE 500E
 CITY: FALLS CHURCH
 STATE: VIRGINIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 22042
 COMPUTER READABLE FORM:


```
FT      repeat_unit      /note= "encodes immunogenic repetitive sequence "
FT      /tag= .453
FT      /tag= "h"
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .480
FT      /tag= .1
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .507
FT      /tag= .1
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .534
FT      /tag= .k
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .561
FT      /tag= .1
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .588
FT      /tag= .m
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .615
FT      /tag= .n
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .642
FT      /tag= .o
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .669
FT      /tag= .p
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .696
FT      /tag= .q
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .723
FT      /tag= .i
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .750
FT      /tag= .s
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .777
FT      /tag= .t
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_unit      /tag= .804
FT      /tag= .u
FT      /note= "encodes immunogenic repetitive sequence "
FT      repeat_region     /note= "encodes immunogenic repetitive sequence "
FT      /tag= .804
FT      /tag= .v
FT      /note= "contains 19 tandem repeats"
XX      US5171843-A.
XX      PD
XX      15-DEC-1992.
XX      PF
XX      30-MAR-1988; 88US-0175112.
XX      PR
XX      12-SEP-1984; 84US-0649903.
XX      PR
XX      09-JUL-1985; 85US-0754645.
XX      PR
XX      26-OCT-1987; 87US-0115634.
XX      PR
XX      30-MAR-1988; 88US-0175112.
XX      PA
XX      (UYNY ) UNIV NEW YORK STATE.
XX      PI
XX      Nussenzeiwg V;
XX      DR
XX      WPI: 1993-008637/01.
XX      DR
XX      P-PSDB; R30609.
XX      PT
XX      New recombinant produced immunogenic poly:peptide - incorporating
XX      PT
XX      part of P. vivax circumsporozoite protein and another part
XX      PT
XX      conserved among different species
XX      PS
XX      Disclosure: Page 14; 18pp; English.
XX      CC
XX      The sequence is that of the Plasmodium vivax circumsporozoite (CS)
XX      CC
XX      gene, the CS protein encoded by the gene contains a repeat
```

```
CC      immunodominant epitope comprising a central region of 19 tandem
CC      repeats of the sequence DRA(D/A)GDPAG. This immunogenic polypeptide
CC      is immunohchemically reactive with monoclonal antibodies against P.
CC      vivax CS protein and is useful in a vaccine against malaria. It may
CC      be prepared by recombinant DNA techniques in yeast.
XX      SQ
XX      Sequence 1137 BP; 393 A; 222 C; 348 G; 174 T; 0 other;
XX
XX      Query Match      3.3%; Score 73.8; DB 14; Length 1137;
XX      Best Local Similarity 58.4%; Pred. No. 6,7e-07;
XX      Matches 129; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
XX
XX      QY      1545 gaatggccagaggaactaaacagccaggttgggagagagatccaagccaggtggg 1604
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      583 gcaagagagatagacagctgtgacagccagcagagagatagagcagcagcagga 642
XX
XX      QY      1605 gccggagagatccaacagccaggtgggctggagagaccacacagccaggtcgccaga 1664
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      643 gataagcagctgtgacacagccagcagagagatagacagatgtgacagcagagagata 702
XX
XX      QY      1665 ggaccacacagcaggttcgggccaagagacccaacagccaggttcggccaagagagacc 1724
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      703 gcagctgtgacagccagcagagagatagacagctgtgacacagcagcagagatagagcagct 762
XX
XX      QY      1725 aaacagccaggttcggccaagagacccaacagccaggttg 1765
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      Db      763 ggaacagcagagagatagacagctgtgacacagcagcagag 803
XX
XX      RESULT      2
XX      N71064
XX      ID      N71064 standard; DNA; 1908 BP.
XX      AC      N71064;
XX      DT      01-JAN-1980 (first entry)
XX      DE      Gene encoding Plasmodium vivax sporozoite circumsporozoite protein.
XX      KW      Immunogen; vaccine; malaria; immunodominant epitope; ss.
XX      OS      Plasmodium vivax.
XX      FT      Key      Location/Qualifiers
XX      CDS      157..1293
XX      /tag= a
XX      /product= circumsporozoite protein
XX      WO8700533-A.
XX      PD
XX      29-JAN-1987.
XX      PF
XX      24-JUN-1986; 86WO-US01373.
XX      PR
XX      12-JUL-1985; 85US-0754645.
XX      PR
XX      (UYNY-) NEW YORK UNIV.
XX      PA      (ARNO/) ARNOT D E.
XX      PI
XX      Arnot DE, Enea V, Nussenzeiwg RS, Nussenzeiwg V;
XX      DR
XX      WPI: 1987-037250/05.
XX      DR
XX      P-PSDB; P70708.
XX      PT
XX      New Plasmodium vivax circumsporozoite protein - and synthetic
XX      PT
XX      peptide(s) contg. its dominant epitope, useful in anti-malarial
XX      PT
XX      vaccines
XX      PS
XX      Disclosure: fig. 3; 32pp; English.
XX      CC
XX      The circumsporozoite protein of P. vivax sporozoites encoded by
XX      CC
XX      this gene is useful in the construction of an anti-malarial
```


XX 11-SEP-1996: 96US-0025929.
 PR (CORR) CORNELL RES FOUND INC.
 PA (UYJE-) UNIV JEFFERSON THOMAS.
 XX
 PI Holloman WK, Kmiec EB, Rice MC, Shu Z, Smith ST;
 XX WPI: 1998-207378/18.
 DR P-PSDB; W56265.
 XX
 PT Mammalian and human REC2 genes encoding ATP-dependent recombinase(s)
 PT - useful for the sensitisation of cells to the apoptotic effects of
 PT irradiation
 PS Claim 12; Fig 1G; 88pp; English.
 XX
 CC This polynucleotide corresponds to the mouse homologue, murec2, of
 CC the REC2 gene (UMREC2) of Ustilago maydis, and codes for an ATP
 CC dependent recombinase (see W56265). Overexpression of mammalian
 CC REC2 in a cell was found to facilitate homologous recombination
 CC between an exogenous DNA and the genome of a cell. It also
 CC sensitises the cell to DNA damage, e.g. by gamma or UV radiation or
 CC from cytotoxic agents. This causes the cell to undergo apoptosis in
 CC response to DNA damage. Murec2 can be used to construct transgenic
 CC animals. REC2 knock-out mice are useful as models for
 CC tumorigenesis induced by chemical carcinogens. The hREC2 human
 CC homologue gene (see V22703) is also claimed, which can be used in
 CC gene therapy methods.
 XX
 SQ Sequence 1525 BP; 450 A; 330 C; 370 G; 375 T; 0 other;
 Query Match 2.7%; Score 60.6; DB 19; Length 1525;
 Best Local Similarity 66.4%; Pred. No. 0.00048;
 Matches 87; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 QY 2120 cccctaaggctcagagggccttcattagctggtgcgaagtgaagcctttct 2179
 DB 1384 cccctaaggagagcagagaggggactaacagattgtaatacagaaggaatttcc 1443
 QY 2180 gaataaaccttttagacttctcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2239
 DB 1444 gaataaagttatgttcgcgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1503
 QY 2240 aaaaaaaaaa 2250
 DB 1504 aaaaaaaaaa 1514
 RESULT 7
 A08554
 ID A08554 standard; DNA; 1525 BP.
 XX
 AC A08554;
 XX
 DT 19-JUL-2000 (first entry)
 DE Murine REC2 serine kinase encoding sequence.
 XX
 KM murec2; serine kinase; homologous pairing; strand transfer; RAD51;
 KM phosphorylation; cyclin E; p53; cell cycle; ss.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT 16..1068
 FT CDS /*tag- a
 FT /product= Rec2
 XX
 PN W0200017329-rl.
 XX
 PD 30-MAR-2000.

XX 17-SEP-1999: 99MO-US21642.
 PE
 XX 21-SEP-1998: 98US-0157603.
 PR
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 PA (CORR) CORNELL RES FOUND INC.
 PA (KIME-) KIMERAGEN INC.
 XX
 PI Havre PA, Rice MC, Holloman WK, Kmiec EB;
 XX WPI: 2000-283562/24.
 DR P-PSDB; Y91933.
 XX
 PT Phosphorylating a serine-containing substrate by incubating it with
 PT adenosine triphosphate and Rec2 kinase and measuring the level of
 PT phosphorylation, useful for discovering specific antagonists or
 PT agonists of Rec2
 PS Disclosure; Page 36; 41pp; English.
 XX
 CC The present sequence encodes the murec2, a murine serine kinase. It is
 CC in the same supergene family as the mammalian protein having homologous
 CC pairing and strand transfer activities, RAD51 and was isolated because of
 CC its homology to the homologous pairing and strand transfer protein of
 CC Ustilago maydis. In particular, murec2 phosphorylates several proteins
 CC that control the cell cycle, especially cyclin E and p53. The invention
 CC permits the phosphorylation of cell cycle control proteins at sites that
 CC are physiologically relevant. The invention can be practiced with either
 CC murine or human Rec2 or a mutant or chimera of these proteins. In
 CC particular the mutant has the sequence of a Rec2 kinase containing other
 CC than a Tyr at position 163. The invention comprises a method of
 CC phosphorylating a serine-containing substrate comprising incubating the
 CC substrate with ATP (adenosine triphosphate) and Rec2 kinase or a
 CC mammalian Rec2 and measuring the level of phosphorylation. The method is
 CC useful for discovering compounds which are specific antagonists or
 CC agonists of Rec2.
 XX
 SQ Sequence 1525 BP; 450 A; 330 C; 370 G; 375 T; 0 other;
 Query Match 2.7%; Score 60.6; DB 21; Length 1525;
 Best Local Similarity 66.4%; Pred. No. 0.00048;
 Matches 87; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 QY 2120 cccctaaggctcagagggccttcattagctggtgcgaagtgaagcctttct 2179
 DB 1384 cccctaaggagagcagaggggactaacagattgtaatacagaaggaatttcc 1443
 QY 2180 gaataaaccttttagacttctcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2239
 DB 1444 gaataaagttatgttcgcgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1503
 QY 2240 aaaaaaaaaa 2250
 DB 1504 aaaaaaaaaa 1514
 RESULT 8
 V55831
 ID V55831 standard; DNA; 799 BP.
 XX
 AC V55831;
 XX
 DT 18-NOV-1998 (first entry)
 DE Nucleotide sequence of the stabilising sequence-encoding insert.
 XX
 KM Fusion protein; stabilising polypeptide; proteolytic degradation;
 KM resistance; half-life; autoimmune disease; inflammation; nitro drug;
 KM IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
 KM nitroreductase protein; enzyme therapy; prodrug therapy; protease;
 KM cancer; pathological condition; ss.

XX OS Epstein-barr virus.
 XX AC X90924;
 XX FN WO9822577-A1.
 XX PD 28-MAY-1998.
 XX PF 17-NOV-1997; 97MO-IB01508.
 XX PR 25-JUN-1997; 97US-0048945.
 XX PR 15-NOV-1996; 96US-0030986.
 XX PA (MASU/) MASUCCI M G.
 XX PI Masucci MG;
 XX DR WPI; 1998-312463/27.
 XX PT New fusion proteins resistant to proteolytic degradation -
 XX PT comprising a core protein with a stabilising polypeptide comprising
 XX PT a peptide sequence containing glycine repeats
 XX PS Disclosure; Fig 4B; 120pp; English.
 XX CC This is a nucleotide sequence of the stabilising sequence-encoding
 XX CC insert. The invention provides a method for increasing the resistance
 XX CC of a core protein to proteolytic degradation that comprises linking or
 XX CC inserting onto or into the core protein a stabilising polypeptide of
 XX CC formula ((GlyA)x(GlyB)y(GlyC)z)n where GlyA, GlyB, GlyC are 1-6
 XX CC sequential Gly residues and x, y, z are Ala, Ser, Val, Ile, Leu, Met,
 XX CC Phe, Pro or Thr and n can be anything between 1-66. x, y and z need not
 XX CC be identical from n repeat to n repeat. Alternatively a nucleic acid
 XX CC encoding the stabilising polypeptide can be linked onto or inserted into
 XX CC a nucleic acid encoding a core protein. The fusion proteins of the
 XX CC invention are more resistant to degradation by proteases and, thus, have
 XX CC a longer half-life than the unfused core protein. The products can be
 XX CC used for treating autoimmune diseases, cancer and inflammation. In
 XX CC particular, the core protein may be an IkappaB regulator protein for the
 XX CC treatment of inflammatory bowel disease, or a nitroductase protein
 XX CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
 XX CC or other pathological conditions. The fusion proteins can also be used in
 XX CC diagnostic methods such as in vivo imaging.
 XX SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;

Query Match 2.7%; Score 60.4; DB 19; Length 799;
 Best Local Similarity 51.9%; Pred. No. 0.00043;
 Matches 136; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 QY 1519 gaggtcctgagggcccaaaagatgggaatggccagaggaactcaaacagcaggttgg 1578
 DB 114 gaggggcccagggagcagagagcagagagggcagagcaggaagagggcagagaggg 173
 QY 1579 gcagggagatcccaaaagcaggtgggcccagagatcccaaacagcaggtgggctggag 1638
 DB 174 caggggggcccagggagcagagaggggcccagagcagggagggcagagaggggagag 233
 QY 1639 gacccaacacagcaggtcgggcccagagaccacaacagcaggtcggcccagagaccaca 1658
 DB 234 cagggaggggggcccagagaggggcccagagcagggagggagggcagagcagagag 293
 QY 1699 aacagcagcaggtcgggcccagagaccacaacagcaggtcggcccagagaccacaacgc 1758
 DB 294 gagggggcccagggagggcagagaggggcccagagaggggcccagagagggcagagagcag 353
 QY 1759 caggtggttggggcccagagcaag 1780
 DB 354 gagggaggggcccagagcagagag 375

RESULT 9
 X90924

ID X90924 standard; DNA; 1925 BP.
 AC X90924;
 DT 17-JAN-2000 (first entry)
 DE Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.
 KW Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid PCWVEBNA; EBNA 1;
 KW episome; transfection; origin of replication; EBV oriP; receptor;
 KW eucaryotic host cell; recombinant cell line; ion channel; gene therapy;
 KW multiple gene expression; transporter protein; transcription factor;
 KW adhesion molecule; antisense therapy; gene amplification;
 KW cell immortalisation; ds.
 XX OS Epstein-barr virus.
 XX FH Key Location/Qualifiers
 XX FT 1..1925
 XX FT CDS
 XX FT /*tag= a
 XX FT /product= "EBNA 1"
 XX FT /transl_except= (pos:799..800, aa:gly)
 XX FT /note= "The sequence is described throughout the
 XX FT specification as being 1926 nucleotides long, but a
 XX FT sequence of only 1925 bp has been given in figure 2"
 XX PN WO9947647-A1.
 XX PD 23-SEP-1999.
 XX PF 12-FEB-1999; 99MO-US03307.
 XX PR 18-MAR-1998; 98US-0040961.
 XX PR 06-AUG-1998; 98US-0130114.
 XX PA (PHAR-) PHARMACOPEDIA INC.
 XX PI Dama] BB, Horlick RA, Robbins AK;
 XX DR WPI; 1999-610610/52.
 XX DR P-PSDB; Y28843.
 XX PT New method for expressing genes from recombinant eukaryotic cells,
 XX PT useful for gene therapy -
 XX PS Claim 24; Fig 2; 86pp; English.
 XX CC The present sequence is a DNA encoding Epstein Barr Virus Nuclear
 XX CC Antigen 1 (EBNA 1), which is obtained from commercially available
 XX CC plasmid PCWVEBNA. EBNA 1 protein is used to stably maintain episomes
 XX CC containing EBV origin of replication (oriP) and a gene encoding
 XX CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
 XX CC protein are transfected with these episomes to produce recombinant
 XX CC cell lines expressing multiple genes of interest. This provides a
 XX CC rapid and reliable method of stably expressing multiple genes in
 XX CC transfected cells. The episomes are useful in the transfection of genes
 XX CC encoding receptors, transporter proteins, ion channels, adhesion
 XX CC molecules and transcription factors. The episomes carrying desired genes
 XX CC can also be used to transfect cells in gene therapy, antisense therapy,
 XX CC for gene amplification, cell immortalisation, etc.
 XX SQ Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;

Query Match 2.7%; Score 60.4; DB 20; Length 1925;
 Best Local Similarity 51.9%; Pred. No. 0.00056;
 Matches 136; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 QY 1519 gaggtcctgagggcccaaaagatgggaatggccagaggaactcaaacagcaggttgg 1578
 DB 356 gaggggcccagggagggcagagaggggcccagagcaggaagaggggcccagagaggg 415
 QY 1579 gcagagatcccaaaagcaggtgggcccagaggaatcccaaacagcaggttgggctggag 1638

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 09:32:22 ; Search time 563.62 Seconds

(Without alignments)
1526.325 Million cell updates/sec

Title: US-09-431-843b-4

Perfect score: 2290
1 tagaattcgcgcgcgcgtga.....cctctcgagtggtggggcgcg 2290

Sequence: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Scoring table: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_36:*

1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	555.6	24.3	789	20	X40044
2	538.2	23.5	802	20	X40045
3	291.2	12.7	375	20	V90541
4	89.4	3.9	114955	20	X53491
5	74	3.2	114955	20	X53491
6	71.2	3.1	1925	20	X90924
7	69.6	3.0	10732	21	A10594
8	68.6	3.0	789	20	X40044
9	67.2	2.9	799	19	V55831
10	67.2	2.9	5452	20	X90923
11	67.2	2.9	8705	20	223778
12	67.2	2.9	9600	19	V21683

13	67.2	2.9	10380	20	Z22248	Nucleotide sequenc
14	67.2	2.9	10596	14	O51731	Plasmid pCISBON f
15	67.2	2.9	10596	17	T40348	Plasmid pCISBON f
16	67.2	2.9	10596	20	X15650	Nucleotide sequenc
17	65.8	2.9	35100	20	V73802	KSHV LTR DNA (nucl
18	65.8	2.9	137507	19	V19941	KSHV long unique c
19	64.8	2.8	1542	10	N90025	DNA encoding anti
20	62.6	2.7	390	13	O21833	Randomising oligon
21	62.6	2.7	390	14	O36859	PCR primer for 5'
22	62.4	2.7	795	19	V55830	FLGA insert stabl
23	61.8	2.7	8438	15	O73500	DNA encoding pseud
24	60.2	2.6	32207	20	V73805	KSHV LTR DNA (nucl
25	60.2	2.6	137507	19	V19941	KSHV long unique c
26	60	2.6	117213	19	V62176	HSV-2 strain SB5 C
27	58.8	2.6	1337	20	Z17263	Human gene express
28	58.4	2.6	150	20	V64956	Mouse histone H2B
29	58.4	2.6	51259	18	X83007	Partial mouse WRN
30	58	2.5	1218	21	A02488	Human colon cancer
31	58	2.5	6225	20	X55273	Human enzyme-relat
32	58	2.5	6225	21	A34721	Human colon cancer
33	57.4	2.5	1593	21	A02504	Human adenosine re
34	56.2	2.5	2277	19	V13836	Human colon cancer
35	56.2	2.5	2277	19	V05372	Homo sapiens mamma
36	56	2.4	1000	21	A02484	Human telomerase p
37	56	2.4	16442	18	X83006	Human colon cancer
38	55.8	2.4	309	10	N90579	Partial mouse WRN
39	55.6	2.4	1091	16	O85844	pDP3.3 cDNA insert
40	54.8	2.4	768	18	T84941	Leishmania donovan
41	54.8	2.4	768	20	X35871	Human prostate pro
42	54.2	2.4	2643	14	O39212	cDNA encoding a pr
43	54	2.4	751	16	O89794	CENP-B cDNA. Homo
44	54	2.4	2169	20	Z07194	Human lung tumour
45	52.6	2.3	1075	15	O64676	Human fibrillarlin

ALIGNMENTS

RESULT 1	
ID: X40044	standard; DNA; 789 BP.
AC: X40044;	
DT: 02-JUL-1999	(first entry)
DE: Prostate cancer associated gene.	
XX: Cancer associated antigen; diagnosis; research; treatment; human;	
KW: breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW: prostate cancer; ss.	
XX: Homo sapiens.	
OS: Homo sapiens.	
XX: MO9904265-A2.	
PN: MO9904265-A2.	
PD: 28-JAN-1999.	
XX: 15-JUL-1998;	98MO-US14679.
PF: 22-JUN-1998;	98US-0102322.
PR: 17-JUL-1997;	97US-0896164.
PR: 10-OCT-1997;	97US-0061599.
PR: 10-OCT-1997;	97US-0061765.
PR: 10-OCT-1997;	97US-0948705.
PR: 11-OCT-1997;	97GB-0021697.
XX: (LUDW-) LUDWIG INST CANCER RES.	
PA: Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;	
XX: Pfeundschnub M, Sahin U, Scanlan MJ, Stockert E;	
PI: Tureci O;	
XX: Tureci O;	

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 XX
 OS Synthetic.
 XX
 PN WO913886-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 17-SEP-1998; 98WO-US19419.
 XX
 PR 09-JUN-1998; 98US-0093972.
 PR 17-SEP-1997; 97US-0059160.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;
 DR WPI; 1999-229400/19.
 XX
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 PS Disclosure: Page 37; 120pp; English.
 XX
 CC The specification describes antisense oligonucleotides (X52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of mRNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of mRNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences X55272-74. These multiple target
 CC oligonucleotides (specifically X55180-271) can be used for the antisense
 CC treatment of diseases and conditions. Typical diseases and conditions
 CC are those associated with impaired respiration and inflammation,
 CC including lung diseases, pulmonary vasoconstriction, inflammation,
 CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
 CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
 CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
 CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
 CC metastases, as well as all types of cancers which may metastasize or have
 CC metastasized to the lungs, including breast and prostate cancer.
 XX
 SO Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 3.9%; Score 89.4; DB 20; Length 114955;
 Best Local Similarity 34.8%; Pred. No. 7,1e-08;
 Matches 537; Conservative 78; Mismatches 913; Indels 14; Gaps 5;

QY 757 gtgcgagggagctgcgaggggtgcgagagctgccttgactacttaattgcgcgtg 816
 Db 106404 GAGCCAGGGGGCCCNHNNNSCGCGCGGCGCCAGCCAGGCGCCCNHNNNSCGCGGC 106345

QY 817 cgcctcgcagaccagcgcgcgcagctgtgtgcacttcgcctctgggaacacttcgcgcgcgc 876
 Db 106344 CGCGGGCGGCGGCGCCCNHNNNSGCGCGGCGCGGGGCGCGAGCCAGGCGCCCN 106285

QY 877 tgcagagctcgtctggggccccaagaacagctcgcgaggttcacagccagctctctgcc 936
 Db 106284 HNNNSAGCGGCGCGCGGCGCGAGCCAGGCGCCCNHNNNSAAACCGCGCGGCGGCGC 106225

QY 937 catcgctcagaggtcctcaggaagtggaagaagaagcccgcgagaccgcgcagcaac 996

Db 106224 GAGCCAGGGGGCCCNHNNNSCAAGCCGGCGCGGCGCGGCGCCAGCCAGGCGCCCNHNNNSG 106165

QY 997 gaggcccaagaccagagctcgaactctgtggccagagcaatacagaggtgtgggagagtg 1056
 Db 106164 CAAGCCGGGCGCGCGGCGCGGCGCGAGCCAGGCGCCCNHNNNSGCAAGCGCGCGGCGGCGC 106105

QY 1057 gacgagggggcccaacacagagcgttgagccccaagagatgtgggaaccttgagagagagc 1116
 Db 106104 GAGCCAGGGGGCCCNHNNNSGGGCGAAGCCGGGCGCGGCGCGAGCCAGGCGCCCNHNNN 106045

QY 1117 cagggggaatgagcagggggccacaggggaagataagcccgagcccttaagcccaagag 1176
 Db 106044 NSCGGCGCAAGCGCGGCGCGGCGCGGCGCGAGCCAGGCGCCCNHNNNSGCGGCGCAAGCGGCGC 105985

QY 1177 agcaagaa-gaagaaagcttgagcttgagccgagcgggagagcgcgcgcacagagagcgcc 1235
 Db 105984 CGCGGGCGCGAGCCAGGCGCGCCCNHNNNSGGCGGCGCAAGCGGCGCGCGGCGCGAGCCA 105925

QY 1236 tcagagtgccctcagaggtggaggaagatcgtctgaatttgaggggtgtgtccctcagcca 1285
 Db 105924 GGGGCGCCCNHNNNSGGGCGGCGAAGCCGGGCGCGGCGCGAGCCAGGCGCCCNHNNNS 105865

QY 1296 gggcagcc-----tcagagcggggaaccacagaaagtgtggtcagagacccttgaggagcag 1351
 Db 105864 CGGCGGCGCAAGCGCGGCGCGGCGCGGCGCGAGCCAGGCGCCCNHNNNSGCGGCGCGCAAG 105805

QY 1352 tcagagctcgcgcgaaccccttgaggagcagaggtgtgcccagaaagtgaagagcggagga 1411
 Db 105804 CGGCGCGCGGCGCGAGCCAGGCGCGCCCNHNNNSGCGGCGCGCAAGCGGCGCGCGG 105745

QY 1412 aggtgatgaggg-----tgcggggagacagtgtcgtgtgccaatgtgtgtggcccaagc 1467
 Db 105744 GCGCGAGCCAGGCGCGCCCNHNNNSGGCGGCGCGGCGCAAGCGGCGCGGCGCGGCGCGCA 105685

QY 1468 ttggcctctgcgggttcctcctgcacacacagagcccaagagcccaagcttgagagagagc 1527
 Db 105684 GGGGCGCCCNHNNNSGGGCGGCGCGGCGCGCAAGCGGCGCGGCGCGGCGCGAGCCAGGCGCCCN 105625

QY 1528 ggggtttggggagacacagaaagtctgaacggggcccaagaagttaccttgagagccca 1587
 Db 105624 HNNNSCGGCGCGGCGCGGCGCAAGCCGGGCGCGGCGCGAGCCAGGCGCCCNHNNNSCGC 105565

QY 1588 tcgagagaccagccagccagccagcagagagcttcagggagagagcagcagcagagagcca 1647
 Db 105564 GCGCGGCGGCGCAAGCGGCGCGGCGCGGCGCGAGCCAGGCGCGCCCNHNNNSGCGGAGCCV 105505

QY 1648 tcgagagaccagccagccagccagccagcagagcttcagggagagagcagcagcagagcca 1707
 Db 105504 GCG-----CCCNHNNNSGCGCGCGCGCCCNHNNNSGCGCGCGCGCCGNHNNNSCGGCGCG 105449

QY 1708 tcgagagaccagccagccagccagccagcagagctcaagggatgagccagccagagccca 1767
 Db 105448 CGGCGNNHNNNSCGGCGCGCGCGCCGNHNNNSCGGCGCGCGCGCGGNHNNNSCGGCGCG 105389

QY 1768 tcgagagaccagccagccagccagccagcagagcttcagggagagagcagcagcagagagcca 1827
 Db 105388 CGGCGCGNNHNNNSCGGCGCGGCGCGGCGCGCCGNHNNNSCGGCGCGCGCGCGGNHNNN 105329

QY 1828 tcgagagaccagccagccagccagccagcagagcttcagggagagagcagcagcagagagcca 1887
 Db 105328 SCGGCGCGGCGCGCGCGCCGNHNNNSCGGCGCGGCGCGCGCGCGGNHNNNSCGGCGCG 105269

QY 1888 tcgagagaccagccagccagccagccagcagagctcaagggatgagccagcagcagagggg 1947
 Db 105268 CGGCGGCGCGCCCNHNNNSCGGCGCGGCGCGCGCGCGCCGNHNNNSCGGCGCGCGCGC 105209

QY 1948 gaagcagcaagagtgagcagkacagagagtgagttcttcagagcttcagggaagccctaa 2007
 Db 105208 GCGCGCGCCCNHNNNSGCGCGCGGCGCGGCGCGCGCCVNNHNNNSGCGCGCGCGCGCGC 105149

QY 2008 g-gaaagagtgcccgctcgagcgtctgtgtcctctgtccctgtcagaggggtctgggagcct 2066
 Db 105148 GCGCGCGCCVGNHNNNSCGGCGCGGCGCGCGCGCGCGCGCCVGNHNNNSGCGCGCGCGCGC 105089

Oy	11313	ggaccaggaatgtggtgcgtctcaggacccctggggagaagcatgtcacgcccttcgcgcacaacccc	1372	
Dd	664	agggcgcaagcaggaaggggcaggaagcaggaaagggcaggaaggaagggcaggaagcagg	743	
Oy	1373	tggggccaggggtggtccgcacaagtgcaggaagcggagagaagctgtgaagggtgtctgggg	1432	
Dd	744	agggcgcaaggaaggggcaggaagcaggaggggcaggaaggggcaggaagcaggagggcagg	803	
Oy	1433	acag 1436 		
Dd	804	gcag 807		
<hr/>				
RESULT	7			
ID	A10594			
XX	A10594 standard; DNA; 10732 BP.			
AC	A10594:			
AT	29-JUN-2000 (first entry)			
DE	Gene encoding a subunit of cellulose synthase.			
KW	Cellulose synthase; cellulose production; increase yield; ds.			
OS	Vigna angularis.			
XX	JP2000060568-A.			
PD	29-FEB-2000.			
PF	26-AUG-1998; 98JP-0239998.			
PR	26-AUG-1998; 98JP-0239998.			
PA	(MIZU/) MIZUNO K. (OUIP) OUI PAPER CO.			
DR	WP1: 2000-342371/30.			
DR	P-PSDB; Y85179.			
PT	A gene encoding a cellulose synthetic equipment - for the improvement			
FT	In the amount of cellulose synthesised in a plant body			
CC	This sequence represents a gene encoding a subunit of the cellulose			
CC	synthase complex of Vigna angularis. The invention relates to subunits of			
CC	cellulose syntehtic equipment, that can be used to increase the amount of			
CC	cellulose synthesised by a plant. The proteins and genes encoding them			
CC	can also be used to improve the properties of the cellulose being			
CC	produced by a plant.			
XX				
SO	Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;			
<hr/>				
Query Match	3.0%; Score 69.6; DB 21; Length 10732;			
Best Local Similarity	13.4%; Pred. No. 0.00022;			
Matches 145; Conservative 441; Mismatches 484; Indels 11; Gaps 4				
Oy	957	gaaggtgaggaaggaaggaagccccggggagcccgacacagagccagagccagaggtgcg	1016	
Dd	9356	glttyargasrghsgargarsyrstrthnasnsrvagsrsraasthrgrtthrtry 9415		
Oy	1017	gaacctgtggccacagacataagcaaggtctggggcaaggttggaagaggggccccagccacg	1076	
Dd	9416	sasnthrtrhgncycysaavaggsaasnysvaygsrasnaaraysasnrgnhsrarvamt 9475		
Oy	1077	gagcgttgagccccaagatgcgggacccttgtagaagagccagggggtgatgagcaggggg	1136	
Dd	9476	hrrartthraasyshsgsrsgnashysgyshsrvaqghnsrtaahvrvaastnsgshrna 9535		
Oy	1137	ccacggggagatatagcccgagccctctaagcccaa-agagagcaagaagaggaacttgg	1195	

Db	9536	asrstrasaaagysraaasylthrsrscysaagsrtrraastrasragsashhargsrsgysys	9595
Qy	1196	agctgacgcgggagggagccgcgcacagacgaagccagccctgaagtgctccaaaggttg	1255
Db	9596	gstrcysgsrsvaagyaasaaasasmntaashvamtgtysaavavsrvgysaaathrgns	9655
Qy	1256	agaagatgcctctgaatttggaggggtgtgcctcaagccagagccctcaagacgggga	1315
Db	9656	rrrsraasrysaavathrgnsnstrgthrgythrsgsrtrhrgggYyqgrvarashsr	9715
Qy	1316	cccaggaagtggcggtgcgcacccctggggagagtcagccctgcgcgcacccc----	1372
Db	9716	gnlthrlthrglthhrvgavaasmrgmtrshshsrassrYsrssasnlhrcysrras	9775
Qy	1373	-tgggagccagggttgccacaaagtgaagagcggaagagtgatgaggttg	1431
Db	9776	hrgytrhvariytrctgnasavavsthrngnashhathrnsysgrrasasasysva	9835
Qy	1432	gacagtgctgcgtgtgccagtggtgtgtgccagaccttgcccttgccggtccctgc	1491
Db	9836	aarghsvaasnashashashsraragysrvahsasnlhrgnasmnrgmcsygs	9895
Qy	1492	ccatgcgggaccccaaggtctgacacagtgagaaaggggttgagagagacaaagat	1551
Db	9896	srcysrrasrrsvraaasnrraagysgnasaaasrrYsargrrgngasrrsgy	9955
Qy	1552	cgaaagcgcccaaaagagtgaccccttgagccccaatcgagaccccaagcccaaccca	1611
Db	9956	sgvaaargr-----aaavaatyrassrrsrssrnysgyarrashsrtrYrasnynmr	10011
Qy	1612	gaaagacctgcaggggagcagccagccagagcccatcgagaccccaagcccccgcgcg	1671
Db	10012	thrgytrmthvargvahgnltrYshrrsgsrYsvagbhsvacystYasarrgystr	10071
Qy	1672	gcaagacctgcaggggagcagccagccagagcccatcgagaccccaagcccccagccg	1731
Db	10072	asltYgvasnhaaYlthrrhaastlrmtashscYshrgnasvasrrmtrargasaay	10131
Qy	1732	gaaggaacca--caagggatgagccagccagagcccatcgagaccccaagcccccgcgc	1789
Db	10132	sysgnastrrasgthraagstrasgygganglthrrsvaagsarghstrargasnstr	10191
Qy	1790	cgagcagacctgcagggagcagcagccagagcccatcgagaccccaagcccccgcgc	1849
Db	10192	gnvahtrhtrsrscygyghsaasnaangytrtytrasngsthlhrasYsaasraags	10251
Qy	1850	cgagcagacctgcagggagcagcagccagccagagcccatcgagaccccaagcccaagc	1909
Db	10252	rgnygsvagargysgnsrathrraasargsmrltasngngsrgyslthassrargns	10311
Qy	1910	cgagcagacctcaagggatgagccagccaaagcgggggagagcagcagaggttcgagka	1969
Db	10312	rhysrsrasngyngnasngysasnngysasnrrasnaagnglnsthrgaagYaaahsas	10371
Qy	1970	cagaggtgagagcttcctgcacagctcgggaagcccttaaggaagagatgccgtcggt	2029
Db	10372	nysvahsrvgggsnysaasrrsnrcnysgggaacycysgngngsmtrtsYsgcysrrsh	10431
DE	Prostate cancer associated gene.		


```
OY 956 ggaaggtgaggaaggaagccccgggacccccacacacagggccagaccacgggtc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 82 ggcagagcagaggaagaggggcagagagggcagagagggcagagagggcagagagc 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1016 ggaacctgtggccagacatagcaagggtg---ggcaggggtggacaggggccccgc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 aggggcagagcagagagaggggcagagaggggcagagaggggcagagaggggc 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1073 cagcagagcgtgagcccccagagatcgcggaacccctggaaagagcagaggggagatggcag 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 aggaacagagagaggggcagagaggggcagagagagagagaggggcagagaggg 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1133 ggggcccaggggaagataggccggagcccttaagcccccaaaagagcagaagaggaagc 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 ggcagagacagagagaggggcagagagcagagagaggggcagagagcagagagc 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1193 tggagctgagccggcgggagcagccgcacacagcagcagccctcagagtgctcagaggg 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 322 aggggcagagaggggcagagaggggcagagagaggggcagagaggggcagaggg 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1253 tggagaagatcgctctgaatttggaggggtgtgcctcagccagcagggcagcctcagaagcg 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 382 aggggcagagaggggcagagaggggcagagagcagagagagaggggcagagggcaggg 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1313 ggaaccacagaggtggcggtcagagacccctggggagcagctgagagccctggccgacacccc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 442 aggggcagagagagagaggggcagagagcagagagaggggcagagaggggcagaggg 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1373 tgggagcagaggttggccacaggtgtaggaagcagagagagaggttggatgaggttctgggg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 502 aggggcagagaggggcagagaggggcagagaggggcagagaggggcagagaggggcaggg 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
X90923/c
ID X90923 standard; DNA; 5452 BP.
XX
AC X90923;
XX
DT 17-JAN-2000 (first entry)
XX
DE Anti-sense strand of pcMVEBNA plasmid.
XX
KW Anti-sense strand; plasmid pcMVEBNA; EBNA 1; episome; transfection;
KW Epstein Barr Virus Nuclear Antigen 1; origin of replication;
KW EBV oriP; eucaryotic host cell; recombinant cell line; ion channel;
KW multiple gene expression; receptor; transporter protein; gene therapy;
KW transcription factor; adhesion molecule; antisense therapy;
KW gene amplification; cell immortalisation; ss.
XX
OS Epstein-barr virus.
OS Cytomegalovirus.
OS Synthetic.
OS
XX Key Location/Qualifiers
XX CDS complement (3032..4957)
XX FT /*tag- a
XX FT /product= "EBNA 1 protein"
XX FT /note= "Epstein Barr Virus Nuclear Antigen 1"
XX
XX W09947647-A1.
XX
XX 23-SEP-1999.
XX
XX 12-FEB-1999; 99WO-US03307.
XX
XX 18-MAR-1998; 98US-0040961.
XX PR 06-AUG-1998; 98US-0130114.
XX
XX (PHAR-) PHARMACOPEDIA INC.
XX
XX DamaJ BB, Horlick RA, Robbins AK;
XX
XX
```

```
DR WPI; 1999-610610/52.
XX
XX New method for expressing genes from recombinant eukaryotic cells,
XX useful for gene therapy.
XX
XX Example 1; Fig 1; 86pp; English.
XX
XX The present sequence is an anti-sense strand of commercially available
XX plasmid pcMVEBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA
XX 1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain
XX episomes containing EBV origin of replication (oriP) and a gene encoding
XX protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
XX cell lines expressing multiple genes of interest. This provides a
XX rapid and reliable method of stably expressing multiple genes in
XX transfected cells. The episomes are useful in the transfection of genes
XX encoding receptors, transporter proteins, ion channels, adhesion
XX molecules and transcription factors. The episomes carrying desired genes
XX can also be used to transfect cells in gene therapy, antisense therapy,
XX for gene amplification, cell immortalisation, etc.
XX
XX Sequence 5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;
XX
XX
XX Query Match 2.9%; Score 67.2; DB 20; Length 5452;
XX Best Local Similarity 47.7%; Pred. No. 0.00055;
XX Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;
XX
XX OY 956 ggaaggtgaggaaggaagccccgggacccccacacacagggccagaccacgggtc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2098 ggcagagcagaggaagaggggcagagaggggcagagaggggcagagaggggcagagagc 2039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1016 ggaacctgtggccagacatagcaagggtg---ggcaggggtggacaggggccccgc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2038 aggggcagagagagagaggggcagagaggggcagagaggggcagagaggggcagaggg 1979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1073 cagcagagcgttggagcccccagagatcgcggaacccctggaaagagcagaggggagatggcag 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1978 aggggcagagagaggggcagagaggggcagagaggggcagagaggggcagaggggcaggg 1919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1133 ggggcccaggggaagataggccggagcccttaagcccccaaaagagcagaagaggaagc 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1918 ggcagagcagagagagaggggcagagaggggcagagaggggcagagaggggcagagaggg 1859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1193 tggagctgagccggcgggagcagccgcacacagcagcagccctcagagtgctcagaggg 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1858 aggggcagagaggggcagagaggggcagagaggggcagagaggggcagaggggcaggg 1799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1253 tggagaagatcgctctgaatttggaggggtgtgcctcagccagcagggcagcctcagaagcg 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1798 aggggcagagaggggcagagaggggcagagaggggcagagaggggcagagaggggcaggg 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1313 ggaaccacagaggtggcggtcagagacccctggggagcagctgagagccctggccgacacccc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1728 aggggcagagagagaggggcagagaggggcagagaggggcagagaggggcagagaggg 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1373 tgggagcagaggttggccacaggtgtaggaagcagagagagaggttggatgaggttctgggg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1678 aggggcagagaggggcagagaggggcagagaggggcagagaggggcagagaggggcaggg 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
223778/c
ID 223778 standard; DNA; 8705 BP.
XX
XX 223778;
XX
XX 14-JAN-2000 (first entry)
XX
XX Vector pShuttle DNA.
XX
XX Antisense; DNA library; identification; multiple cloning site; MCS;
XX inhibition; ss.
XX
XX
```



```
KM proteolysis resistant; liver; malignancy; CMV-driven;
KM Cytoomegalovirus; episomal expression plasmid; ss.
XX Synthetic.
XX Key
FH enhancer
FT location/Qualifiers
FT 1..611
FT /tag= a
FT /note= "CMV enhancer/promoter"
FT 758..775
FT /tag= b
FT /label= SP6_promoter
FT 845..849
FT /tag= c
FT /note= "SP6 RNA start"
FT 902..966
FT /tag= d
FT /function= cloning_linker
FT 967..1107
FT /tag= e
FT /note= "SV40 poly A"
FT 1108..1131
FT /tag= f
FT /function= SV40_origin
FT 1580..4189
FT /tag= g
FT /label= EBNA-1
FT 4190..6374
FT /tag= h
FT /function= oriP
FT 4295..4887
FT /tag= i
FT /note= "family of repeats"
FT 5866..5978
FT /tag= j
FT /note= "dyad region"
FT 6375..6457
FT /tag= k
FT /label= HSV_TK_terminator_3'-end
FT 6975..7975
FT /tag= l
FT /phenotype= neomycin_resistance
FT /note= "Tn5 neomycin phosphotransferase gene"
FT 7975..8112
FT /tag= m
FT /label= TK_promoter
FT 8114..8594
FT /tag= n
FT /function= M13_orf1
FT 8595..10414
FT /tag= o
FT /label= delta_2a
XX
XX W09323541-A.
XX
XX 25-NOV-1993.
XX
XX 17-MAY-1993; 93WO-US04648.
XX
XX 18-MAY-1992; 92US-0884811.
XX
XX 18-MAY-1992; 92US-0885971.
XX
XX (GETH ) GENENTECH INC.
XX
XX Godowski PJ, Lokker NA, Mark MR.
XX
XX WPI; 1993-386573/48.
XX
XX Hepatocyte growth factor variants - are resistant to proteolytic
XX cleavage into its two-chain form, used to treat malignancies
XX associated with HGF receptor
XX
XX Example 1; Fig 6; 87pp; English.
PS
```

```
XX Plasmid pCISBON (a PRK5 derivative) is an episomal CMV driven
CC expression plasmid. HuHGF variants with enhanced receptor binding
CC activity were produced by site-directed mutagenesis. Stable
CC populations of preferred HGF variants were obtained by transfecting
CC human embryonic kidney 293 cells and then these were subcloned in
CC pCISBON. See R52940-R52949 for examples of pref. HGF variants.
XX
SQ Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T; 0 other;

Query Match 2.9%; Score 67.2; DB 14; Length 10596;
Best Local Similarity 47.7%; Pred. No. 0.00038;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

OY 956 ggaagctgagagaaagaaagcccgaggagcccgcaagagagccagccagcgtc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2287 ggcagagagcagagagagagagagagagagagagagagagagagagagag 2346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1016 ggaacctgtggccagagcatagcaaggtgg--ggcagaggtggacgagggcccaagc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2347 agggcagagagcagagagagagagagagagagagagagagagagagagagag 2406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1073 cagcagagcgtgagcccgagatcgagcccgagagagagagagagagagagagag 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2407 aggaagcagagagagagagagagagagagagagagagagagagagagagagag 2466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1133 ggggccaagggagagatagcgcggagcccttaagccccaagagagcaagagagaaac 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2467 ggcagagagcagagagagagagagagagagagagagagagagagagagagagag 2526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1193 tggagcttgagccgagagagagagagagagagagagagagagagagagagagag 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2527 aggggcagagagagagagagagagagagagagagagagagagagagagagag 2586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1253 tggagagagatcgctctgaattggaggggtgtgtcccccagccagagagagagagcgg 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2587 aggggcagagagagagagagagagagagagagagagagagagagagagagag 2646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1313 ggaaccagagaaagtggcggtcagagacccttgggagagcagtcagccctgcgcaacc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2647 aggggcagagagagagagagagagagagagagagagagagagagagagagag 2706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1373 tgggagccagaggtgagccagacaggtgagagagcgagagagaggtggtatgagaggtg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 aggggcagagaggggcagagagagagagagagagagagagagagagagagagag 2766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
T40348
ID T40348 standard; DNA; 10596 BP.
XX
XX T40348;
XX
XX 09-DEC-1996 (first entry)
XX
XX Plasmid pCISBON for expression of hepatocyte growth factor.
XX
XX Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
XX pro-hormone; beta subunit; alpha subunit; kringte domain; prothrombin;
XX plasmidogen; catalytic domain; serine protease; HGF variant;
XX HGF receptor; malignancy; chronic HGF receptor activation; ss.
XX
XX Synthetic.
XX
XX US5547856-A.
XX
XX 20-AUG-1996.
XX
XX 18-MAY-1992; 92US-0884811.
XX
XX 13-JUL-1993; 93US-0087783.
XX
XX 18-MAY-1992; 92US-0884811.
PR
```


Search completed: October 12, 2000, 17:29:33
Job time: 28631 sec

```
PR 18-MAY-1992; 9205-0885971.
XX
PA (GETH ) GENENTECH INC.
XX
PI Godowski PJ, Loker NA, Mark MR;
XX
DR WPI; 1996-392634/39.
XX
PT New hepatocyte growth factor variants - are resistant to in vivo
XX
PT proteolytic cleavage into a 2-chain form, useful as HGF antagonists
XX
PS
XX
PS Example 1; Fig 6; 39pp; English.
CC
CC This sequence represents the episomal CMV driven expression plasmid
CC pcisBON which was used in the expression of variant human hepatocyte
CC growth factor (HGF). HGF is isolated from human serum and is a
CC disulphide linked heterodimer derived by proteolytic cleavage of the
CC pro-hormone between residues 494 and 495. This generates a molecule
CC composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and
CC a beta subunit of 234 amino acids (mol. wt. 34 kD). The alpha and beta
CC subunits are encoded by a single open reading frame. The alpha subunit
CC contains four kringle domains based on their homology to kringle-like
CC domains in other proteins, e.g. prothrombin, plasminogen. The beta
CC subunit shows high homology to the catalytic domain of serine proteases.
CC However two of the three residues which form the catalytic triad of
CC serine proteases are not conserved in HGF. Therefore, the precise
CC function of the beta chain remains unknown. The invention includes HGF
CC variants which retain HGF receptor binding activity without having the
CC biological activity of wild-type HGF. They can be used for the treatment
CC of pathological conditions associated with the activation of a HGF
CC receptor such as malignancies associated with chronic HGF receptor
CC activation. The pcisBON plasmid comprising the HGF coding sequence
CC may be used for manipulation of the HGF coding sequence and expression
CC of the variant HGF's of the invention.
XX
SQ Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T; 0 other;

Query Match 2.98; Score 67.2; DB 17; Length 10596;
Best Local Similarity 47.7%; Pred. NO. 0.00058;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggttgaggaaggaaggaagcccgagcccgagccagcagcagccaggttc 1015
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2287 ggcagagcaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2346
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1016 ggaactgtggccagagcatagcaaggtgg---ggcagagtgagcagagggccagc 1072
   |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 2347 aggggcagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggggc 2406
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1073 cagcagacgttgagagcccaagatcggggaacccctgagagaaggaaggaagga 1132
   |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 2407 agggagcaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 2466
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1133 ggggcccaggggaagataagcggcggcccttaagccccaagagaagcaagagaagc 1192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2467 ggcagagcaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 2526
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1193 tggagctgagccagcagcagcagcagccacagagcagagccctcagagtgctccacagag 1252
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 2527 aggggcagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2586
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1253 tggagaagatgcctcgaatttgagaggggtgtgcctcagcagggcagagcctcagagcag 1312
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 2587 aggggcagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2646
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1313 ggaaccaggaagtgaggcgtcagagcccttgaggaggaagtgacgcccctggccgccaacccc 1372
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2647 aggggcagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2706
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1373 tgggagcagaggttgccagacaaggtgagagaagcagaggaaggtgatatgaggtgtctgggg 1432
   ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 2707 aggggcagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 2766
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 05:20:12 ; Search time 7878.83 Seconds
(without alignments)
1797.047 Million cell updates/sec

Title: US-09-431-843b-4
Perfect score: 2290
Sequence: 1 tagaattcgcgcgcgcctga.....ccttcgagtgaggggcgcg 2290

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 segs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gsa1:*
89: gb_gsa2:*
90: gb_gsa3:*
91: gb_gsa4:*
92: em_gsa1:*
93: em_gsa2:*
94: em_gsa3:*
95: em_gsa4:*
96: gb_gsa15:*
97: gb_gsa16:*
98: gb_gsa17:*
99: gb_gsa18:*
100: gb_gsa19:*
101: em_gsa5:*
102: em_gsa6:*
103: em_gsa7:*
104: em_gsa8:*
105: em_gsa9:*
106: em_gsa10:*
107: em_gsa11:*
108: em_gsa12:*
109: em_gsa13:*
110: gb_gsa5:*
111: gb_gsa6:*
112: gb_gsa7:*
113: gb_gsa8:*
114: gb_gsa9:*
115: gb_gsa10:*
116: gb_gsa11:*

117: gb-ss812:*
118: gb-ss813:*
119: gb-ss814:*
120: gb-ss820:*
121: gb-ss821:*
122: gb-ss822:*
123: gb-ss823:*
124: gb-ss824:*
125: em-ss814:*
126: em-ss815:*
127: em-ss816:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635	27.7	729	34	BE264806 601192867
2	596.6	26.1	768	34	BE277227 601178592
3	592.6	25.9	751	35	BE387193 601277069
4	586	25.6	682	35	BE385629 601278145
5	515	22.5	606	21	AM338856 ha67f02.x
6	514	22.4	522	14	AL120919
7	509	22.2	526	24	AM732857
8	506.2	22.1	591	35	BE410591 601303311
9	476.2	20.8	493	35	BE336852 ba97e12.y
10	466.2	20.4	551	34	BE207411 601290328
11	455.2	19.9	529	35	BE398108 601290328
12	454.6	19.8	550	21	AM249563 2821312.3
13	454.2	19.8	546	21	AM246100 2821312.5
14	454.6	19.4	1296	35	BE409569 601299883
15	435.4	19.0	757	35	BE391750 601283311
16	425.8	18.6	433	10	A1459806 ap17g11.x
17	423	18.5	432	13	AT816189 aa45h07.x
18	418.6	18.3	526	22	AM513841 x049h03.x
19	416.8	18.2	502	34	BE206098 ba97e12.x
20	415	18.1	432	21	AM250708 2822396.5
21	413.4	18.1	427	34	BE302499 ba66d10.y
22	400.8	17.5	654	34	BE282484 601103075
23	398	17.4	549	35	BE391975 601285002
24	392.2	17.1	400	13	AT816230 aa45h07.y
25	381	16.6	473	19	AM072475 x06d12.x
26	375.4	16.4	886	34	BE274933 601123054
27	366.8	16.0	454	13	A1926535 wa6bd12.x
28	360.6	15.7	445	11	A1590782 tw25d09.x
29	355.8	15.5	451	12	AT634247 tr44b12.x
30	354.4	15.5	590	8	AT115047 u14f17.y
31	354.8	15.5	735	36	BE534864 601231408
32	348	15.3	436	10	A1434006 th33d03.x
33	339.4	14.8	564	9	A1173869 ue92f06.y
34	332.4	14.5	559	23	AM631605 90800 MAR
35	306.4	13.4	414	21	AM326593 19380 MAR
36	299.2	13.1	500	24	AM762644 ur64h02.y
37	297.8	13.0	496	35	BE489073 167730 BA
38	291.2	12.7	506	21	AM321684 u037c01.y
39	287.2	12.5	379	36	BE504832 h233a03.x
40	278.8	12.2	554	25	AM966375 ESR375648
41	274.4	12.0	1050	34	BE288040 601094726
42	259.8	11.3	447	13	A1786854 u332d09.y
43	254.4	11.1	256	36	BE551638 h396d08.x
44	251	11.0	363	37	H31247
45	244	10.7	355	7	AA865212 ch28e05.s

ALIGNMENTS

RESULT	1
BE264806	
944	tcgagggctccaggaaggtgaggaaggaagccccgggagcccccagaccaggaagcca 1003

Db 421 TCGAGGCTCCAGAGGTGAGAGAGAAAGACCCCGGGAGACCCGACGAGGCCA 480
Qy 1004 gaccccaaggtctgagcctgttgccagagcatagcaaggtctggagcaggttgacg-ag 1062
Db 481 GCACCCAGAGGTCTGTGGCCAGACATAGCAAGGGTGGGGCAGGGTGAGCGAG 540
Qy 1063 gggcccaagcagaggtctgagcctgagcagcagccttgagagagagcagag 1122
Db 541 GGGCCCAAGCAGAGGTCTGTGGCCAGACATAGCAAGGGTGGGGCAGGGTGAGCGAG 599
Qy 1123 gatgagcagaggggagcaggggagagatagggcagccttaagcccaagagagcaag 1182
Db 600 GATGAGCAGAGGGCCAGCGGAGATAGAGCCGAGCCCTTAAGCCCAAGAGAGCAAG 659
Qy 1183 aagaggaagctgtagctgagcggcgagagcagcccgcccaag 1225
Db 660 -AGAGGAAGCTGGAGCTGCCGGGGAGCAGCCCAAGAGCAG 701

RESULT 2
LOCUS BE277227 868 bp mRNA EST 13-JUL-2000
DEFINITION 601178592F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3050950 5',
mRNA sequence.
ACCESSION BE277227
VERSION BE277227.1 GI:9152194
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 868)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
Plate: LNCM3 row: h column: 23
High quality sequence stop: 742.

FEATURES
source Location/Qualifiers
1..868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3050950"
/clone_id="NIH_MGC_20"
/tissue_type="melanocytic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5-
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-DNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 171 a 245 c 294 g 157 t 1 others
ORIGIN

Query Match 26.1%; Score 596.6; DB 34; Length 868;
Best Local Similarity 88.1%; Pred. No. 1.1e-111;
Matches 760; Conservative 0; Mismatches 74; Indels 29; Gaps 9;
Qy 44 ccgactcgactccactctgagagagagagatcgagagacgcgagagcagag 103
|||||

Db 1 CCGAGCTGCGACTCCACTGTGGAGAGAGAGAGATGCGGAGAGCGGAGAGCAGAG 60
Qy 104 actgc-gagagagcgagagc-----gccggcgagagagcgagc 145
Db 61 ACTGCTGAGGAGCGGGATGGGCTTGCCTTGCCTGAGGTGACTGTGTCTCAGAGGTAC 120
Qy 146 cagggagcagagagagagtcgagagagcgcgagcgcgagcccgagcctcttcagt 205
Db 121 TGAGACCATGATGTGCGATGATGATGCTGCTGCTGCGGCTCGGCCAGCTGTTCCAGT 180
Qy 206 ccagaatgacagaggtccagaaac-tggcgagccacgagagacatgtgtatgagac 264
Db 181 CCAGATATGACAGGGGTCAGAACTTGCGAGCCACAGAGGACATAGTATGATGCGCAC 240
Qy 265 aactatccgagatctgttgagacgagactgcaatgggagacagccaaactgaatttcac 324
Db 241 AACTATCCGGATCTGTGTGAAGAGACTGCAATGGGGACAGCCAAACTGATTCTAC 300
Qy 325 agaattgagtcgcgcttcctcccaagcgctgtttcatgtgagagacatttcagaactgg 384
Db 301 AGAATGAGATCCGCTTCTCTCCCAAGCGCTGTTTCAATTGAGACATTCTCAGAACTGG 360
Qy 385 acggaacactatgacctcttgagagacaatcactctacalcagtggtgttctctctg 444
Db 361 ACGGACACATATGACCTCTTGAGGACAAATCATCTCATTCAGATGCTGCTTCTCTG 420
Qy 445 cggagacagagagtgaaatgcatgccaagccctcaagctcagagaggtctgaggttt 504
Db 421 CGAGAACGAGAGTGAATGTCATGCTCAAGCCCTCAGCTCAGAGGAGTGAAGTGT 480
Qy 505 aaaaatcccaagagagatccagagcgaggtgtccgggagcctcagagctatgctggcttc 564
Db 481 AAAAGCTCCAGAGATCCAGAGAGCGGCTTCCGGGCTTCAGAGCTATCTTGCGCTTC 540
Qy 565 tac-ggagtcgcgctgagagaccagagcagagcaggttgaggcgagagacagaaatcca 623
Db 541 TACGGGAGATCCGGCTGAGGAGACCGAGGACCGGAGCGGTGGGCGGAGCAGAACTACCA 600
Qy 624 gaagcgcttccagaaactgaaactgagcagcacaacaacttcgcatcaagcattcct 683
Db 601 GAAGGCTTCCAGAACCTGAACTGAGCGAGCAGCAACAACTCCGACACAGCATCTCT 660
Qy 684 caagtcgcgctgagcagcagccttcagcagcagcagcagcagcagcagcagcagcagc 741
Db 661 CAGTCTGCTGCTGAGTACCTGCGCTCTGACACTTCCAGGCGCCCGCTGCTCTTC 720
Qy 742 ctgga-ggagacgctgtgtcg--cgagagctgtccgggg--tgcgagagagtgccctgg 796
Db 721 CTGGAGGAGAGAGCGTGGTGGGGGGGTGAGCTGGCGGGGGGTGGCGGAGAGTGCTCG 780
Qy 797 --actactcatgttcgctgtcgctgcgacacacagcgcgcagctgtgtgacattgc 854
Db 781 GACTTACTTCAATGTGCGCGCTGCGCTGCGACACAGAGCTGCAGAGTGTTGCCCTCGC 840
Qy 855 ctggagagcattccggcccgct 877
Db 841 CTGGGAGCTTTCGGGCGCGGT 863

RESULT 3
LOCUS BE387193 751 bp mRNA EST 21-JUL-2000
DEFINITION 601277069F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3618136 5',
mRNA sequence.
ACCESSION BE387193
VERSION BE387193.1 GI:9332558
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by BMF (Biomedical Research Center at the Charité,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.

No sl sequence available.
This clone (DKFZ762B143) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
Location/Qualifiers

source
1..522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZ762B143"
/clone_lib="762 (synonym: hmel2)"
/tissue_type="melanoma (Memo cell line)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 121 a 155 c 148 g 98 t
ORIGIN

Query Match 22.4%; Score 514; DB 14; Length 522;
Best Local Similarity 99.0%; Pred. No. 7.2e-95;
Matches 517; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
OY 221 ccagaactgagcgacagagagacatgtgtatgcgcacacatccgacatctg 280
Db 1 CCAGAAATCGGAGCCACGAGGACATGTAGTATCGGACAACTATCCGATCTGG 60
OY 281 tggagcagactgacatgaggagacagcacaacatgagttctacagaaatagatccgct 340
Db 61 TGGAAACGAGACTCCAAATGGGACGACGCCAAACCTGAGTTCTACAGAAATAGATCCGCT 120
OY 341 tctgcccacagcgtctgtcttctatgagacatctctcagaactgagcagacatagacc 400
Db 121 TCTGCGCCACGGGCTGTTCATGTAGGACATTTCTCAAGACTGGACGCAACTATAGCC 180
OY 401 tcccttgaggaaatcattccctacatccagctgctgttcctctcgagagacaggaatga 460
Db 181 TCCCTGAGGAATCTACATCTCAATCCATCCAGTGGCTGTTCCCTCGGAAACACAGAGTGA 240
OY 461 actgcatgccaagccccctcaccagctcagagggagctgagagtgcttaaaaagctccagagaga 520
Db 241 ACTGGCATGCCAAGCCCTCAACGCTCAGAGAGGTGAGGTCTTTAAAAAGCTCCCAAGAGA 300
OY 521 tccagagagcgtctgtccgggacctagagatcagctgctgtctctacaggaatccgagctg 580
Db 301 TCCAGGAGCGGCTTGTCTCGGCGCTACGAGCTCATGCTGGCTCTACGCGGATCCGGCTGG 360
OY 581 agggccagagagagcgagcggtggggccgagacagaactacaaagcgcttccagaacc 640
Db 361 AGGACCGAGAGACGGGACGGGTGGGCGAGGACAGAACTACCAAAAGGCTTCCAGAAACC 420
OY 641 tgaactggcgagcagcacaacaacacctccgcatcacacgcatcctcaagtgcgctgtagc 700
Db 421 TGAAGTGGCGGACGACACAACACTCCGCAATCACACGCAATCTCAAGTCCGTGGGTGAGC 480
OY 701 tggagcctcgagcaattcccaagcgccaactggtccgctcttc 742
Db 481 TGGGCTTCGAGCACTTCCAGCGCGGCTTGGTCCGCTCTTTC 522
```

RESULT 7
AW732857 526 bp mRNA EST 21-APR-2000
LOCUS
DEFINITION bdl5g10.y1 NIH_MGC-21 Homo sapiens cDNA clone IMAGE:2963010 5'
similarity to TR:096029 096029 7-60. ;, mRNA sequence.
ACCESSION AW732857
VERSION AW732857.1 GI:7633196
KEYWORDS EST.

SOURCE

ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ARCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://imga.llnl.gov/image/html/resources.shtml>
Seq primer: -40RP from Gibco
High quality sequence stop: 515.

FEATURES

Location/Qualifiers

source
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2963010"
/clone_lib="NIH_MGC-21"
/tissue_type="Choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGG(6). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 108 a 168 c 193 g 57 t
ORIGIN

Query Match 22.2%; Score 509; DB 24; Length 526;
Best Local Similarity 98.1%; Pred. No. 7.5e-94;
Matches 515; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
OY 1216 ccggccacagagcagagccctcagagtgctcagaggttgagagaatcgctgaattg 1275
Db 1 CCGCCCATTTGCCAGGCGCTCTCAGAGTGGCTCTCAGAGTGGAGAAATCGCTGAAATTGG 60
OY 1276 gagggtgtgctccctcagcagcgacgctcagaagcggagacccaggaagtggcgctcag 1335
Db 61 GAAGGGGTGTCCCTCAAGCCAGGCGACCTCAGAGCGGGGACCCAGAAATGGGGGTCCAG 120
OY 1336 gaccctggggagagcagtgcaagccctgcgcacaaaccccttggagcgaggttggccagaag 1395
Db 121 GACCCTGGGGAGGACAGTGCAGCCCTCCGCCAACCCCTGGAGACCGAGGTGGCGCAAG 180
OY 1396 gtaaggaagggaggaagtgtgagtggaggtgtcgggggaagaagtgctgagtgccagtggt 1455
Db 181 GTAGGAAAGCGGAGAAAGGAGTGAAGGAGTGGGAGAGTGTCTGGGTGGCAATGGT 240
OY 1456 ggtgccacaaccttggcccttgcgggtcccttgcgcccatcggggaaccccaaggtctga 1515
Db 241 GGTGCCACAGACTTGGCCCTTGGCGGGTCCCTGCCCATCGGGGAGACCCCAAGGTGGA 300
OY 1516 caaatgagaaggggttgaagagacacagaagtgcaacgggggcccacaagaagtacc 1575
Db 301 CACAGTGAAGAACGGGCTTAGGAGGACACATAAGGTCAACGGGGCCCAAAAGGATACC 360
OY 1576 cctggagagccatcggagagccccaagggccccaagcagagagccttgcagggagagcca 1635
Db 361 CCTGGAGAGCCCATCGAGAGCCCAAGGCCCCCGCCCAAGCAAGCTGCAAGGGAGCAAGCCA 420
OY 1636 gccagagaccatcgaagagccccaagccccgcgcggcagagaccttgcagggagagacca 1695
```


Db 421 GCCGAGAGCCCATCGAGAGACCCCATGCCCCGCGAGCGAGCATACAGGATGAGCCA 480
 QY 1656 gccgagagcccatcgagagaccacccagcccgccgagacct 1740
 Db 481 GCCGAGAGCCCATCGAGAGACCCCATGCCCCGCGAGCGAGCAT 525

RESULT 8

LOCUS BE410591 591 bp mRNA EST 21-JUL-2000
 DEFINITION 601303311F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637634 5',
 mRNA sequence.
 ACCESSION BE410591
 VERSION BE410591.1 GI:9347041
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
 Plate: LNCM337 row: f column: 03
 High quality sequence start: 95
 High quality sequence stop: 591.

FEATURES

source 1..591
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3637634"
 /clone_id="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 136 a 149 c 196 g 109 t 1 others

Query Match 22.1%; Score 506.2; DB 35; Length 591;
 Best Local Similarity 99.4%; Pred. No. 2.8e-93;
 Matches 508; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 55 tccaccctggagagagacagagagatgagagagcgagagagagagagagagagagagc 114
 Db 81 TCACATGTGTGAGAGAGACAGAGATGCGAGAGACCCGAGAGACAGAGACTGCCAGGAC 140
 QY 115 gccgagagcccgccgag 174
 Db 141 GGCAGAGCGCGCGCGAGGAGAGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 200
 QY 175 ccgag 234
 Db 201 CCGCGGCA 260
 QY 235 gccag 294

Db 261 GCCAGAGAGGAGATGTAGATGATGGCACAACACTTCGGAGATCTGGAGAGAGACTGC 320
 QY 295 aatggag 354
 Db 321 AATGGAG 380
 QY 355 ttttcattgag 414
 Db 381 TGTTCATTGAGAGAGATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 440
 QY 415 cactccatcatcag 474
 Db 441 CACTCCTCATCATCATGAG 500
 QY 475 cccctcag 534
 Db 501 CCCCTCAG 560
 QY 535 gtcgag 591
 Db 561 GTCCGGGCTTCAG 591

RESULT 9

LOCUS BE336852 493 bp mRNA EST 14-JUL-2000
 DEFINITION ba37e12.y1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:2957902 5',
 similar to TR:096029 096029 7-60. ;, mRNA sequence.
 ACCESSION BE336852
 VERSION BE336852.1 GI:9189237
 KEYWORDS EST.
 SOURCE human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Other_ESTs: ba37e12.x1
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 image.lnl.gov/image/html/lresources.shtml
 Seq primer: -40RP from Glibco
 High quality sequence stop: 423.

FEATURES

source 1..493
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2957902"
 /clone_id="NIH_MGC_14"
 /tissue_type="renal cell adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: kidney; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 103 a 150 c 186 g 53 t 1 others
 ORIGIN


```

/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:3629948"
/clone_1lb="NIH_MGC_21"
/tissue="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 327 a 416 c 389 g 164 t

ORIGIN

Query Match 19.4%; Score 443.6; DB 35; Length 1296;
Best Local Similarity 84.7%; Pred. No. 1.6e-80;
Matches 669; Conservative 0; Mismatches 89; Indels 32; Gaps 14;

QY 515 aggaatccagagc-ggctgtccggcctacagacatgctgtggtcttcaagatc 573
|||||
Db 11 AGGAGATCCAGAGCTACTGTCCGGGCTACAGCTCATGCTGGGCTTACGGATC 70

QY 574 cggctgagagacagagcagg-gcaaggtggccgagacagaactaccagaagcctt 632
|||||
Db 71 CGGCTGGAGGACCGACGCGCGCGCGCGCGAGCAGACAGAACTACGAAACGCTT 130

QY 633 ccaagactgagctggcagcagcacaacaactccgcatcaacagcatcgaatggcc 692
|||||
Db 131 CCAGACTTGAACTGCGCAGCAGCAGCAACCTCCGATCACACCATCTTCAATGCT 190

QY 693 gtgtgagctgagcctcagacacttccagcgcacatgtgtccgtcttctcctggaagac 752
|||||
Db 191 GGGTGAAGCTGGGCTCGAGCACTTCCAGCGCGCTGGTCCGCTTCTTCTGGAAGAC 250

QY 753 gctgtgctggcgagagctcgccgggggtcgagagatgctccttggaacttctt-c 811
|||||
Db 251 GCTGTGGGGGGGAGCTGCGGGGGTGGGAGAGTCCCTGGAATCTTCAATTTTCA 310

QY 812 ccgtgagctgagcagacagc-gccgcagcgtgtgtcacttgccttggaagacttccg 870
|||||
Db 311 CCGTGCCTGCGCAGCAGCAGCAGCAGCAGCTGTGCTTCTGAGAGCACTTCCG 370

QY 871 cccgcg-tgcaagtcgt--ctgggggcccagaagaagcttgagggttcaagccagc 927
|||||
Db 371 ACCGCAATGCAAGTTCGTAATGGGGGCCACAGAACAGCTGCGGAGTTCAAGCCAGC 430

QY 928 tctctgcccat-cgctcagaggtctcagaagaagtggagaagaagcccgagg 986
|||||
Db 431 TCTTGCCCAATACCGCTGAGGGCTCCAGGAAGCTGAGAGGAAGCCCGGGGA 490

QY 987 ccccgacagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1040
|||||
Db 491 CCCCACCAAGAGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 550

QY 1041 ggtgtgggagaggtgtgagcgaaggggcccagcagcagcagcagcagcagcagcagc 1097
|||||
Db 551 GCGGGGCAAGAGTGGACAGAGGAGCCAGGAGCGTGGAGGCCCAAGATGCCG 610

QY 1098 gggagccctggagagagcagcaggggga---tgagcagagggggcagcaggggaagttagc 1154
|||||
Db 611 GGAGCAATGCAAGAGAGCAGCGGGAGCTGAGCGCAGCGGGGCCCAAGCAATCA 670

QY 1155 ggaag-----cccttaagcccaagaagcagaagaagagagcagcagcagcagcagc 1204
|||||
Db 671 GAGGAGCCCATATACGACGAAACAGAGAGCAGCAGCAAGAAAGATGGCCCTGACCC 730

QY 1205 ggcggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1264
|||||
Db 731 CGCGAGAGCATGCC--CCACAGAGCCCAAGGCTTCAAGTGGCTCA-AGGTGCAAGAGCAGC 788

QY 1265 ctctgaattt 1274
Db 789 CACTGAAGTT 798

RESULT 15
LOCUS BE391750 757 bp mRNA EST 21-JUL-2000
DEFINITION 601283311F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3604994 5',
mRNA sequence.
ACCESSION BE391750
VERSION BE391750.1 GI:9337115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L10W252 row: f column: 03
High quality sequence start: 110
High quality sequence stop: 686.
Location/Qualifiers
1. 757
source

FEATURES
source

BASE COUNT 170 a 204 c 251 g 131 t 1 others

ORIGIN

Query Match 19.0%; Score 435.4; DB 35; Length 757;
Best Local Similarity 92.9%; Pred. No. 7.2e-79;
Matches 577; Conservative 0; Mismatches 31; Indels 13; Gaps 11;

QY 51 cgaatccacccgggagagagagagagagcagcagcagcagcagcagcagcagcagcagc 110
|||||
Db 94 CAATTCAACTGGGAGAGAGAGAGAGATGCGGATACCGGAGAGAGAGAGATGCGGA 153

QY 111 ggaagcagagccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 170
|||||
Db 154 GGAGGCGAGAGCGCGCGCGCGAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGTCCGA 213

QY 171 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 229
|||||
Db 214 GGAAGCGCGGCGGAGCGGCTCCATCTGCTTCAGATCCAGATAGAGGTCAGAAACT 273

QY 230 ggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 287
|||||
Db 274 GGCGAGCAGCAGAGAGAGATGTGTACGATGTGGCAGCAACTATCCGGATTCGTGGGAAGC 333

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 17:29:33 ; Search time 563.62 seconds
(without alignments)
1604.975 Million cell updates/sec

Title: US-09-431-843b-5

Perfect score: 2408
Sequence: 1 tagaatcagcgccgcctga.....ccctctgagtggtggggggcg 2408

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_36: *
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT: *
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT: *
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT: *
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT: *
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT: *
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT: *
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT: *
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT: *
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT: *
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT: *
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT: *
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT: *
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT: *
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT: *
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT: *
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT: *
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT: *
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT: *
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT: *
21: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT: *
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612.6	25.4	789	20	X40044
2	547.8	22.7	802	20	X40045
3	302	12.5	375	20	V90541
4	91.2	3.8	114955	20	X53491
5	85.2	3.5	114955	20	X53491
6	80.8	3.4	1542	10	N90025
7	75.8	3.1	789	20	X40044
8	72.4	3.0	10732	21	A10594
9	71.2	3.0	1925	20	X90924
10	68.2	2.8	35100	20	V73802
11	68.2	2.8	137507	19	V19941
12	67.2	2.8	799	19	V55831

C	13	67.2	2.8	5452	20	X90923
C	14	67.2	2.8	8705	20	Z23778
C	15	67.2	2.8	9600	19	V21683
C	16	67.2	2.8	10360	20	Z22248
C	17	67.2	2.8	10596	17	Q51731
C	18	67.2	2.8	10596	17	T40348
C	19	67.2	2.8	10596	20	X15650
C	20	64	2.7	795	19	V55830
C	21	62.6	2.6	390	13	Q21833
C	22	62.6	2.6	350	14	Q36859
C	23	61.8	2.6	8438	15	Q73500
C	24	61.2	2.5	1218	21	A02488
C	25	60.4	2.5	1091	16	O85844
C	26	60.2	2.5	32207	20	V73805
C	27	60.2	2.5	137507	19	V19941
C	28	60	2.5	117213	19	V62176
C	29	58.8	2.4	761	16	O89794
C	30	58.8	2.4	1337	20	Z17263
C	31	58.4	2.4	150	20	V64956
C	32	58.4	2.4	51259	18	X83007
C	33	57.4	2.4	600	6	N50853
C	34	57.2	2.4	6225	20	X55273
C	35	57.2	2.4	6225	21	A34721
C	36	56.2	2.3	2277	19	V13836
C	37	56.2	2.3	2277	19	V05372
C	38	55	2.3	16442	18	X83006
C	39	55.8	2.3	309	10	N90579
C	40	55.4	2.3	1000	21	A02484
C	41	55.4	2.3	1593	21	A02504
C	42	54.8	2.3	768	18	T84941
C	43	54.8	2.3	768	20	X35871
C	44	54.2	2.3	2643	14	Q39212
C	45	54	2.2	2109	20	Z07194

ALIGNMENTS

RESULT 1
X40044 standard; DNA; 789 BP.
XX
AC X40044;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
(LUDW-) LUDWIG INST CANCER RES.
PA
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX

Anti-sense strand
Vector pshuttle DN
Vector plasmid pcm
Nucleotide sequenc
Plasmid pcisEBON f
Plasmid pcisEBON f
Nucleotide sequenc
Fick insert stabl
Randomising oligon
PCR primer for 5'
DNA encoding pseud
Human colon cancer
Leishmania donovan
KSHV LTR DNA (nucl
KSHV long unique C
HSV-2 strain SB5 C
Leishmania donovan
Human gene express
Mouse histone H2B
Partial mouse WRN
Sequence encoding
Human enzyme-relat
Human adenosine re
Homo sapiens mamma
Human telomerase p
Partial mouse WRN
pD9.3 cDNA insert
Human colon cancer
Human prostate pro
cDNA encoding a pr
CENR-B cDNA. Homo
Human lung tumour

DE		Human adenosine A1 receptor antisense oligonucleotide fragment.
XX		
KW		Antisense oligonucleotide; multiple target; antisense treatment;
KM		impaired respiration; inflammation; lung disease;
KW		pulmonary vasoconstriction; inflammation; allergic rhinitis;
KV		acute asthma; allergy; asthma; impaired respiration;
KW		respiratory distress syndrome; pain; cystic fibrosis;
KV		pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW		chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW		colon cancer; breast cancer; lung cancer; pancreatic cancer;
KM		hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW		prostate cancer; ss.
XX		
OS	Synthetic.	
XX		
PN	WO913886-A1.	
XX		
PD	25-MAR-1999.	
XX		
PF	17-SEP-1998; 98WO-US19419.	
XX		
PR	09-JUN-1998; 98US-0093972.	
XX		
PR	17-SEP-1997; 97US-0059160.	
XX		
PA	(UYEC-) UNIV EAST CAROLINA.	
XX		
NY	Nyce JW;	
PI		
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary	
XX	vasoconstriction	
XX		
DR	WPI: 1999-229400/19.	
XX		
PS	Disclosure; Page 37; 120pp; English.	
XX		
CC	The specification describes antisense oligonucleotides (X52869-X55271)	
CC	directed against at least 2 mRNAs selected from target genes, coding and	
CC	non-coding regions of RNAs corresponding to target genes, gene	
CC	initiation codons, genomic flanking regions, intron-exon borders, the	
CC	5'-end, the 3'-end and the juxta-section between coding and non-coding	
CC	regions and all segments of RNAs encoding proteins associated with one	
CC	or more diseases, conditions or mixtures. The antisense oligonucleotides	
CC	may be derived from sequences X5572-74. These multiple target	
CC	oligonucleotides (specifically X53180-271) can be used for the antisense	
CC	treatment of diseases and conditions. Typical diseases and conditions	
CC	are those associated with impaired respiration and inflammation,	
CC	including lung diseases, pulmonary vasoconstriction, inflammation,	
CC	allergic rhinitis, acute asthma, allergies, asthma, impaired respiration,	
CC	respiratory distress syndrome, pain, cystic fibrosis, pulmonary	
CC	hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive	
CC	pulmonary disease (COPD), and cancers such as leukemias, lymphomas,	
CC	carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic	
CC	cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic	
CC	metastases, as well as all types of cancers which may metastasize or have	
CC	metastasized to the lungs, including breast and prostate cancer.	
XX		
SO	Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other:	
	Query Match	3.8%; Score 91.2; DB 20; Length 114955;
	Best Local Similarity	31.7%; Pred. No. 3.3e-08;
	Matches 534; Conservative 139; Mismatches 994; Indels 15; Gaps 5;	
OY	697 gagctgaacccctcgagcaattcaggcggccaccatgcgttcgcttcttcctgagagaagactg 756	
DG	104123 gggccggggcsmndnngcgtcggcggcggsnmndncgtcggcggcgsgcnmdnngtgc 104182	
OY	757 gtctgcgcggagactgcaggggggtctgcggcagagtgcctctgactactcatgtctgcgcgtg 816	
DG	104183 gggcgcggcgscnnndnttgggcggcgsgcnmdnngggcggcgsgcnmdnncgcgbgccc 104242	
OY	817 cgcgtgcgacacacgcgccacagcttgtcaccttcgccttcggtagaaccttcgcggccccgc 876	

[illegible]

PF 12-FEB-1999; 99WO-US03307.
XX
PR 18-MAR-1998; 98US-0040961.
PR 06-AUG-1998; 98US-0130114.
XX
PA (PHAR-) PHARMACOPEIA INC.
XX
PI DamaJ BB, Horlick RA, Robblins AK;
XX
XX WPI: 1999-610610/52.
DR P-PSDB; Y28843.
XX
PT New method for expressing genes from recombinant eukaryotic cells,
XX useful for gene therapy -
XX
XX Claim 24; Fig 2; 86pp; English.
XX
XX The present sequence is a DNA encoding Epstein Barr Virus Nuclear
CC Antigen 1 (EBNA 1), which is obtained from commercially available
CC plasmid PCMVEBNA. EBNA 1 protein is used to stably maintain episomes
CC containing EBV origin of replication (oriP) and a gene encoding
CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
CC protein are transfected with these episomes to produce recombinant
CC cell lines expressing multiple genes of interest. This provides a
CC rapid and reliable method of stably expressing multiple genes in
CC transfected cells. The episomes are useful in the transfection of genes
CC encoding receptors, transporter proteins, ion channels, adhesion
CC molecules and transcription factors. The episomes carrying desired genes
CC can also be used to transfect cells in gene therapy, antisense therapy,
CC for gene amplification, cell immortalisation, etc.
XX
XX Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;

Query Match 3.0%; Score 71.2; DB 20; Length 1925;
Best Local Similarity 48.1%; Pred. No. 9.6e-05;
Matches 233; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggtgagaggaagaagcccgaggaccgacacagggccagaccgggtc 1015
DB 334 ggcagagcagagagagagggcagagagggcagagagggcagagagggcagagag 383
QY 1016 ggaacctggaaccagacatagcagaggtg--ggcaggggtgacagagggcccgac 1072
DB 384 agggcagagagcagagagagagggcagagggcagagagggcagagagggcagaggg 443
QY 1073 caccgagcgtgagaccacagatgcggagacccttgagagagagcagagggagtgagcag 1132
DB 444 aggaagcagagagagggcagagggcagagcagagagagagggcagagagggcagaggg 503
QY 1133 ggggccaaggggaagataggcccgagacccttaagcccaagaagacagagaggaagc 1192
DB 504 ggcagagacagagagagggcagagcagagagggcagagagggcagagggcagaggg 563
QY 1193 tggagctagccggcggcagcagcccgccacagggcagggccctcagaggtgctcaggag 1252
DB 564 agggcagagagggcagagagggcagagcagagagagggcagagagggcagagggcag 623
QY 1253 tggagaagatcgctctgaatttgaggggtgtgacctcagccagcagggcagcctcagagag 1312
DB 624 agggcagagagggcagagcagagggcagagggcagagcagagagggcagagggcagag 683
QY 1313 ggaaccagaagatggcgtcaggaacctgggagagcagtgtagacccctgcggcacaacc 1372
DB 684 agggcagagagagggcagagggcagagcagagagggcagagagggcagagggcagag 743
QY 1373 tgggcccagaggtggtggccacaggtgtaggaagcagagaggtgtaggtactggag 1432
DB 744 agggcagagagggcagagagggcagagagggcagagagggcagagagggcagagggg 803
QY 1433 acag 1436
DB 804 gcaag 807

RESULT 10
ID V73802 standard; DNA; 35100 BP.
XX V73802;
AC V73802;
XX
DT 25-FEB-1999 (first entry)
XX
DE KSHV LUR DNA (nucleotides 1-35,100).
XX
KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW diagnosis; treatment; BHV8; complement binding protein; v-cBP; SSBP;
KW sDNA binding protein; transport protein; glycoprotein B; pol; vIL-6;
KW DNA polymerase; viral interleukin-6; BHV4-IE1 I; thymidylate synthase;
KW vMIP-II; BHV4-IE1 II; vMIP-I; capsid protein I; tegument protein I; ds.
XX
OS Kaposi's sarcoma-associated herpesvirus.
XX
PN US5849564-A.
XX
PD 15-DEC-1998.
XX
PE 29-NOV-1996; 96US-0770379.
XX
PR 29-NOV-1996; 96US-0770379.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JT;
XX
XX WPI: 1999-069741/06.
XX
PT Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT dihydrofolate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
XX
XX Disclosure; Column 67-96; 109pp; English.

CC This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for KI, ORF4 which encodes the complement binding protein
CC v-cBP, ORF6 which encodes a sDNA binding protein (SSBP), ORF7 which
CC encodes a transport protein, ORF8 which encodes glycoprotein B, ORF9
CC which encodes DNA polymerase (pol), ORF10, ORF11, K2 which encodes viral
CC interleukin-6 (vIL-6), ORF02 which encodes dihydrofolate reductase
CC (DHFR), K3 which encodes BHV4-IE1 I, ORF70 which encodes thymidylate
CC synthase, K4 which encodes vMIP-II, K5 which encodes BHV4-IE1 II, K6
CC which encodes vMIP-I, K7, ORF16 which encodes Bcl-2, ORF17 which encodes
CC capsid protein I, ORF18 and ORF19 which encodes Tegument protein I.
CC KSHV is a new human Herpesvirus (HHV8) believed to cause Kaposi's sarcoma
CC (KS) which is the most common form of neoplasm occurring in persons with
CC acquired immune deficiency syndrome (AIDS). The DHFR protein is useful
CC for vaccination, prophylaxis, diagnosis and treatment of a subject with
CC Kaposi's sarcoma and for detecting expression of a DNA virus associated
CC with Kaposi's sarcoma in a cell.
XX
XX Sequence 35100 BP; 8703 A; 9395 C; 8921 G; 8081 T; 0 other;

Query Match 2.8%; Score 68.2; DB 20; Length 35100;
Best Local Similarity 48.2%; Pred. No. 0.00042;
Matches 227; Conservative 0; Mismatches 238; Indels 6; Gaps 1;

QY 1593 gaccacagggcccgccagcagagagcctgcaggggagagcaccacagaccatcgga 1652
DB 24307 gagcccgccgagcaccacccagagaccccgccgagcaccacccagcagcaccac 24366
QY 1653 gaccacagggcccgcccgccagagcctacagaggtatgagccagcagagcccatcgga 1712
DB 24367 gagcccgccgagcaccacccagagcctcccgccgagcaccacccagagcaccac 24426

```

QY 1713 gacccagagcccccgcgcagagactgtcagggagacagccagagcccatcgga 1772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24427 gagccccgcagacaccccaagaccccgagaccccaagagcccgagaccccaag 24486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1773 gacccagagcccccgcgcagagactgtcagggagacagccagagatcccatcgga 1832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24487 gagccccgcagacaccccaagagaccccgagaccccaagagcccgagaccccaag 24546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1833 gacccagagcccccgcgcagagactgtcagggagacagccagagagcccatcgga 1892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24547 gagccccgcagacaccccaagagaccccgagaccccaagagcccgagaccccaag 24600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1893 gacccagagcccccgcgcagagactgtcagggagacagccagagagcccatcgga 1952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24601 ccccaagagcccccgcgcagacccctcccgaggggagatcccgagcccaacccccc 24660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1953 gacccagagcccccgcgcagagactgtcagggagacagccagagagcccatcgga 2012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24661 gagggggagatcccgagccaccccccggaggggagatcccgagccacccccc 24720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2013 gacccagagcccccgcgcagagactgtcagggagacagccagagc 2063
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24721 gagggggagatcccgagccaccccccggaggggagatcccgagccgc 24771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
V19941
ID V19941 standard; DNA; 137507 BP.
XX
AC V19941;
XX
DT 03-AUG-1998 (first entry)
XX
DE KSHV long unique coding region and terminal repeat.
XX
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW complement-binding protein; glycoprotein; capsid protein IV; infection;
KW immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW lymphoproliferative disease; leukemia; splenomegaly; mycosis fungoides;
KW HIV immune status; anti-inflammatory agent; therapy; ds.
XX
OS Kaposi's sarcoma-associated herpes virus.
XX
XX
FH Location/Qualifiers
FT 1142..2794
FT /tag= a
FT /product= complement-binding protein
FT 8699..11236
FT /tag= b
FT /product= glycoprotein B
FT complement (117261..17875)
FT /tag= c
FT /product= interleukin 6
FT complement (21548..21832)
FT /tag= d
FT /product= macrophage inflammatory protein II
FT complement (27137..27424)
FT /tag= e
FT /product= interferon regulatory factor I
FT 2861..29741
FT /tag= f
FT /product= protein TI.1
FT complement (58976..60175)
FT /tag= g
FT /product= glycoprotein M
FT complement (69412..69915)
FT /tag= h
FT /product= glycoprotein L
FT complement (88410..88910)
FT /tag= i
FT /product= interferon regulatory factor 2

```

```

FT CDS
FT 89600..90541
FT /tag= j
FT /product= interferon regulatory factor 3
FT 90173..90643
FT /tag= k
FT /product= glycoprotein X
FT complement (93636..94127)
FT /tag= l
FT /product= interferon regulatory factor 4
FT complement (111931..112443)
FT /tag= m
FT /product= capsid protein IV
FT complement (123808..127296)
FT /tag= n
FT /product= immediate early protein

MO9804576-A1.
PD 05-FEB-1998.
XX
XX 22-JUL-1997; 97WO-US13346.
PF
XX 29-NOV-1996; 96US-0757669.
PR 25-JUL-1996; 96US-0686243.
PR 25-JUL-1996; 96US-0686349.
PR 25-JUL-1996; 96US-0686350.
PR 25-JUL-1996; 96US-0687253.
PR 25-JUL-1996; 96US-0688814.
PR 05-SEP-1996; 96US-0708678.
PR 10-OCT-1996; 96US-0728323.
PR 13-NOV-1996; 96US-0747887.
PR 13-NOV-1996; 96US-0748640.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
PA
XX
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo J;
XX WPI: 1998-130615/12.
XX
XX New nucleic acid encoding Kaposi's sarcoma associated herpes virus
FT proteins - useful for, e.g. detecting levels of HHV8 in, and
FT preparation of vaccines for treatment of, HIV patients
PT
XX
XX Example 2; Page 135-203; 230pp; English.
XX
XX This sequence represents the long unique region and terminal repeat of
XX the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
XX as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
XX invention which encode KSHV polypeptides selected from: (a) viral
XX macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
XX (c) viral Irf 1; (d) complement-binding protein; glycoproteins B, M or L;
XX (d) capsid protein IV encoded by ORF5; and (e) immediate early protein
XX encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
XX by it, and antibodies (Ab) specific for the proteins are useful for
XX detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
XX fluids or tissue samples. HHV8 infections can be treated with antisense
XX or triplex forming molecules or agents that bind specifically to the
XX protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
XX while the protein can be used in protective vaccines. Ab may also be used
XX to differentiate between lymphomas, and HHV8 may be implicated in many
XX other lymphoproliferative diseases such as lymphomas, leukaemia,
XX splenomegaly and mycosis fungoides. Cells and animals containing the
XX nucleic acid are useful for drug screening. HHV8-derived peptides can be
XX used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
XX can be inhibited with methotrexate. These can also be used to determine
XX the immune status of a patient infected with HIV. HHV8 derived protein
XX viral MIP III may be used as an anti-inflammatory agent for,
XX e.g. treating rheumatoid arthritis. This sequence is stated as containing
XX 81 open reading frames.
XX
SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

```



```
Oy 1016 ggaactgtgagccagagcatagcaaggttg---ggcagaggttgacagagggcccaagc 1072
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7903 AGGGCAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7844
Oy 1073 caggaagcgttgagcccaagagatgcggagaccccttgagagagagagagagagagag 1132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7843 AGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7784
Oy 1133 ggggacagagagagagagagagagagagagagagagagagagagagagagagagagc 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7783 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7724
Oy 1193 tgaagctgagccgagagagagagagagagagagagagagagagagagagagagagag 1252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7723 AGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7664
Oy 1253 tggagaagatcgctctgaatttgagaggggtgtgcctcagagagagagagagagagc 1312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7663 AGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7604
Oy 1313 ggaaccagagagagagagagagagagagagagagagagagagagagagagagagagc 1372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7603 AGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7544
Oy 1373 tggagagcagaggtggccgacaaagtgaggaagcggagagagagagagagagagagag 1432
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7543 AGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7484
```

```
RESULT 15
V21683
ID V21683 standard; DNA; 9600 BP.
```

```
AC V21683;
```

```
DT 17-AUG-1998 (first entry)
```

```
DE Vector plasmid pCMVkmITR-EPI.
```

```
KM Polynucleotide delivery; plasmid pCMVkmITR-EPI; vector;
```

```
KW gene therapy; vaccine; polycationic agent; ss.
```

```
XX Chimeric - Epstein-Barr virus.
```

```
OS Chimeric - Adeno-associated virus.
```

```
OS Chimeric - Cytomegalovirus.
```

```
OS Chimeric - Bos taurus.
```

```
Key Location/Qualifiers
```

```
FT CDS
```

```
FT 14..2594
```

```
FT /tag= a "EBV nuclear antigen A"
```

```
FT misc-feature
```

```
FT 2623..4559
```

```
FT /tag= b "EBV origin of replication"
```

```
FT repeat_unit
```

```
FT 4928..5104
```

```
FT /tag= c "AAV inverted terminal repeat"
```

```
FT repeat_unit
```

```
FT 7189..7335
```

```
FT /tag= d "AAV inverted terminal repeat"
```

```
FT promoter
```

```
FT 5112..6734
```

```
FT /tag= e "CMV immediate-early enhancer/promoter"
```

```
FT terminator
```

```
FT 6818..7050
```

```
FT /tag= f "note= 'bovine growth hormone polyA sequence'"
```

```
XX WO9806437-A2.
```

```
XX 19-FEB-1998.
```

```
PF 13-AUG-1997; 97WO-US14465.
XX
XX 13-AUG-1996; 96US-0023867.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Cohen F., Dubois-Stringfellow N., Dwarik V., Innis MA;
XX Murphy JE, Tetsuo U, Zukermann R;
XX
XX WPI: 1998-159296/14.
```

Polycationic agents based on alpha-amino acids, able to complex

with nucleic acid - to facilitate its entry into cell, condense it

and protect it against serum degradation, particularly for use in

gene therapy

Disclosure; Page 77-80; 100pp; English.

This polynucleotide comprises the DNA sequence of vector plasmid

pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin

of replication from plasmid pCEP4, a coding region for EBV nuclear

antigen A from pCEP4, a pair of inverted terminal repeats from

adeno-associated virus, a cytomegalovirus enhancer/promoter, a

bovine growth hormone polyA sequence, and a kanamycin resistance

selectable marker. Polynucleotides encoding polypeptides, such as

erythropoietin or leptin, and ribozymes and antisense

polynucleotides can be inserted into the vector. The vector is

preferred for use in novel compositions and methods for improved

polynucleotide delivery into cells. In these methods, polycationic

agents are used to increase the frequency of uptake of a

nucleic acid (see also V21684-86) into a cell. The polycationic

agent can condense with the nucleic acid and inhibit serum and/or

nuclease degradation of the nucleic acid. The nucleic acid can be

a vector, may express a therapeutic protein or a vaccinating viral

or cancer antigen, or is itself therapeutic (antisense or

ribozyme). The methods and compositions can be used in the gene

therapy of many diseases.

Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 2.8%; Score 67.2; DB 19; Length 9600;

Best Local Similarity 47.7%; Pred. No. 0.00058;

Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

```
Oy 956 ggaaggttgagagagagagagagagagagagagagagagagagagagagagagagag 1015
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 ggcagagagagagagagagagagagagagagagagagagagagagagagagagagag 812
Oy 1016 ggaactgtgagccagagatgcagaggttg---ggcagaggttgagagagagagagagc 1072
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 agggagagagagagagagagagagagagagagagagagagagagagagagagagag 872
Oy 1073 caggaagcgttgagcccaagagatgcggagaccccttgagagagagagagagagagag 1132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 agggagagagagagagagagagagagagagagagagagagagagagagagagagag 932
Oy 1133 ggggacagagagagagagagagagagagagagagagagagagagagagagagagagc 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 ggcagagagagagagagagagagagagagagagagagagagagagagagagagagag 992
Oy 1193 tggagctgagccgagagagagagagagagagagagagagagagagagagagagagag 1252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 agggagagagagagagagagagagagagagagagagagagagagagagagagagag 1052
Oy 1253 tggagaagatcgctctgaatttgagaggggtgtgcctcagagagagagagagagagc 1312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1053 agggagagagagagagagagagagagagagagagagagagagagagagagagagag 1112
Oy 1313 ggaaccagagagagagagagagagagagagagagagagagagagagagagagagagc 1372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 agggagagagagagagagagagagagagagagagagagagagagagagagagagag 1172
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 12:02:09 ; Search time 7878.83 seconds
(without alignments)
1889.646 Million cell updates/sec

Title: US-09-431-843b-5
Perfect score: 2408
Sequence: 1 tagaattcagcgccgcctga.....cccttcgtgagtggtggggcg 2408

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss15:*
97: gb_gss16:*
98: gb_gss17:*
99: gb_gss18:*
100: gb_gss19:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: em_gss12:*
109: em_gss13:*
110: gb_gss5:*
111: gb_gss6:*
112: gb_gss7:*
113: gb_gss8:*
114: gb_gss9:*
115: gb_gss10:*
116: gb_gss11:*

|||||
Db 421 TCAGAGGCTCCAGGAAGTGTGAGAGAGGAAGCCCCGGGAGACCCGACACAGAGCCA 480
QY 1004 gcaacccaggtctgagctgtgagccagacatgaagagtggtggcagagtgagc-ag 1062
Db 481 GCACCCAGGCTGGACTGTGGGCGCACAGCATACCAAGGGTGGGGGCTGTGACCAAG 540
QY 1063 gggcccccagcagcagcgtgagagcccccagatgcggagcccttgagagagagcagggg 1122
Db 541 GGGCCCCAGCACGAGACGTGGAGCCCGAGATGCGGAGCCCTGTGAGAGAGAGCCA-GGG 599
QY 1123 gatgagcagggggccacggggaagatagagccggagcccttaagcccccaagagagcag 1182
Db 600 GATGAGGACAGGGGGCCACGGGGAAGATAGCGCCGAGCCCTTAAGCCCAAGAAGAGCAG 659
QY 1183 aagaagagctgtgagctgtgagccggcgagcagcagccagcag 1225
Db 660 -AGAGGAACTGGAGCTGGCCGGGGAGCAGCGCCAAAGCAG 701
RESULT 2
BE277227 868 bp mRNA EST 13-JUL-2000
LOCUS 601178592F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3050950 5',
DEFINITION mRNA sequence.
ACCESSION BE277227
VERSION BE277227.1 GI:9152194
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 868)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph. D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DC/DIP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
Plate: LNC97 row: h column: 23
High quality sequence stop: 742.
Location/Qualifiers
1..868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3050950"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(s). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 171 a 245 c 294 g 157 t 1 others
ORIGIN
Query Match 24.8%; Score 596.6; DB 34; Length 868;
Best Local Similarity 88.1%; Pred. No.5.6e-111;
Matches 760; Conservative 0; Mismatches 74; Indels 29; Gaps 9;
QY 44 ccgactcgactccactgtgagagagcagagagagtgagagcagcagagcagcagcag 103
|||||

Db 1 CCGACTGCACTCCACTGTGGAGAGAGCAGAGAGATGCCGAGAGACCGGAGAGAGG 60
QY 104 actgc-gagagcggcgagggc-----gcggcgcgagggagcgagc 145
Db 61 ACTGCTAGACGCGCGATGCTGCTGCTGCGCTGAGGTGACTGTGTACTGCAAGGGTAC 120
QY 146 caggggcagagcagagagtcgagagagcggcgaggcgagggccagctcttcagt 205
Db 121 TGAAGACGATGAGATGTGGATGTGATGCTCTGTGGCGCTGTGGCCCTGCTTCCAGT 180
QY 206 ccaaatgacagaggtccagaagac-tggcgagccacgagagagacatgtgtagatcagcag 264
Db 181 CCAAAATGACAGGGTCCAGAAACTTGGCGAGCCACGAGGAGACATGTGTAGTATCGGCAC 240
QY 265 aactatccgagctgtgtgagacgagactgcaatgtgggacagccaaactgagttctac 324
Db 241 AACTATCCGATCTGTGGAGAACGAGATGCAATGGGGACACGGCCAAACCTGAAATTTCTAC 300
QY 325 agaaatgagatccgcttcctccgccaagcgctgttctatgtagacatcttcagagactgg 384
Db 301 AGAATAGATTCGCTTCTGCTCCCAAGGCTGTTCATTAGAGCATCTTCCGAACCTGG 360
QY 385 acgacacactatgacctccttgaggaacatcactcctacatccagtgtgttctctg 444
Db 361 ACGGACAACTATGACCTCCTTGAGGAAATCACTCCTACATCCAGTGGCTGTTCTCTG 420
QY 445 cgagaaacagagatgaaactgtgcatgccaagccctcaagctcagggaggtcagagtgtt 504
Db 421 CGAGAACCAAGAGATGAATGTCATGTCACCAAGCCCTCACGCTCAGGAGGTGAGGTGTT 480
QY 505 aaaaagctccaggaagttccagagcgagctgtctcgagccctcagagctcatgtcgggttc 564
Db 481 AAAAGCTCCAGAGATCCAGAGCGGCTGTCTCGGGCTTACAGACTCATGTGGCTTC 540
QY 565 tac-ggagtcggcgctgagagacagcagcagcggtggtggcgagacagactacca 623
Db 541 TAGGGGATCCGGCTGTGAGGACCGAGGCAAGGGCAGCGTGGCGAGCAAGAACTACCA 600
QY 624 gaagcgcttcagaaacctgaactgtgcgacgacacaacaacctcgcacacgcatcct 663
Db 601 GAAGCGCTTCCAAACCTGAACCTGAGCGCACACCAACCAACCTCCGATCACAGCATCCT 660
QY 684 caagtcggcgctgtgagtgagctgagcacttcagcag--cgcaactgtgcctcttc 741
Db 661 CAACTGCTGGGTGAGTGGGCTCGAGCACTTCAGGGCGCGCTGCGCTCTCTTC 720
QY 742 ctgga-ggaagcgtggtcgsg--cggaagctgcggggg--tgcgagagatgacctg 796
Db 721 CTGAGGGAACCCCTGTGCTGGGGCGGTGAGACTGCCGGGGGTGCCGCAAGGTGCTGG 780
QY 797 -actactcaatgtcgcgctgtgcgtgcgacacacagcgcgcagctgtgacctgcg 854
Db 781 GACTTACTTATGATGTCGCGCTGCGCTGCCGACACACAGGCTGCCAGTGTGCTCCGCG 840
QY 855 ctgggagcacttcggggccgct 877
Db 841 CTTGGAGGCTTTCGGGCGCGGT 863
RESULT 3
BE387193 751 bp mRNA EST 21-JUL-2000
LOCUS 601277069F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3618136 5',
DEFINITION mRNA sequence.
ACCESSION BE387193
VERSION BE387193.1 GI:9332558
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DC/DMP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: Image.lnl.gov
Plate: LNCM286 row: 1 column: 17
High quality sequence start: 86
High quality sequence stop: 727.
Location/Qualifiers
1..751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3618136"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 170 a 202 c 254 g 125 t
ORIGIN

Query Match 24.6%; Score 592.6; DB 35; Length 751;
Best Local Similarity 96.8%; Pred. No. 3.6e-110;
Matches 626; Conservative 0; Mismatches 19; Indels 2; Gaps 2;
QY 10 ggggcgcgtgaattctagccgagcatgagaccccgactcgactcaacttggagag 69
DB 38 GCGAGGCGCGGATGTCGTACTATGTGAGACTCGACTCCACCTCGGAGAG 97
QY 70 gacgagagatgacgagagacgagagac-gaggaactgagagagagccgagccg 128
DB 98 GACGAGAGAGATCGGAGAGACCGGAGAGCTGCGAGAGCGGAGCGCGG 157
QY 129 cgcgagagacgagcagcagagagagagagagtcgagagagccgagcgcg 188
DB 158 CCGAGAGGACGCGGAGCGAGGAGAGAGAGTCGAGAGCGCGGCGGCG 217
QY 169 gccagagctgttccagatcagaatgacagaaactgagagccagagagacat 248
DB 218 GCCAGGCTGTTCCAGATGACAGAGGTCAGAACTGCGCAGCAGAGGACAT 277
QY 249 gtgtagatcgcgacaactatccgagatcgttgagacagagactcaatggagac 308
DB 278 GGTAGGTATCGGACACTATCCGAGTGTGTGAGACGAGACTCATGGGAGAC 337
QY 309 aaacctgagttctacagaatgagatccgcttcctgcccacagggcttcaatg 368
DB 338 AAACCTGAGTTTCTACAGAAATGAGATCCGCTTCCTGCCAAACGGCTTTCA 397
QY 359 catcttgaagactgagagacaactatgactccttgaagacaactactctaatca 428
DB 398 CATTTTCAAGACTGAGAGGACACTATGACTCTTGAAGCAATCACTCTATATCA 457
QY 429 gtgtgctgttctcctcgcagaaacagagtgagctgcatgccaagccctcaagct 488
DB 458 GTGGCTGTTTCTCTGCGAAGACAGAGTGAATCGCATGCGCAGGCCCTCAG 517
QY 489 gtagagtcagagtggttaaaagctccagagagatccagagcggttgcgagctaca 548
DB 518 GAGAGTCAGAGTGTTTAAAGCTCCACAGAGATCCAGAGCGGCGTTGTCCGGCT 577

QY 549 gctcatgctggctcttctacagggatcccgctggagagaccagagcgagcagct- 607
DB 578 GCTCATGCTGGCTTCTACGGGATCCGCTGGAGACCGAGCGGCGACGGGTGG 637
QY 608 gtagcagaactaccagaagcgcttccagaactgaaactgagcagc 654
DB 638 GAGCAGCAAGACTACAGAGAGCGCTTCCAGAACTGAACTGAGCGCAGC 684
RESULT 4
BE385629 682 bp mRNA EST 21-JUL-2000
LOCUS BE385629
DEFINITION 601278145F1 NIH_MGC_20 Homo sapiens CDNA clone IMAGE:3618922 5', mRNA sequence.
ACCESSION BE385629
VERSION BE385629.1 GI:9330994
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
ARTHOBS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DC/DMP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: Image.lnl.gov
Plate: LNCM288 row: j column: 11
High quality sequence start: 20
High quality sequence stop: 653.
Location/Qualifiers
1..682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3618922"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 160 a 190 c 224 g 108 t
ORIGIN

Query Match 24.3%; Score 586; DB 35; Length 682;
Best Local Similarity 96.2%; Pred. No. 7.7e-109;
Matches 654; Conservative 0; Mismatches 20; Indels 6; Gaps 5;
QY 40 gacccgactgagctccac-tgggaagagagagagagatgcgagagcagagaga 98
DB 3 GACCCGAGTGTGAGTCCACCATGAGAGGAGCGAGGAGATGCCGAGACCCGAG 62
QY 99 cgaagagtcgagagc 158
DB 63 CGAGAGATCGAGAGAGCGGAGAGCCGCCGCCGCGAGAGACCGAGGAGCAG 122
QY 159 cgaagagtcgagagc 218
DB 123 CGAGAGATCGAGAGAGCGGAGAGCCGCCGCCGCGAGACTGTTCCAGATGAC 182

QY	219	gtccgaacacgagcgagccacagaggaacatgtgtgagatcgcgaacatctccgacct	278
Db	183	gtccgaacacgagcgagccacagaggaacatgtgtgagatcgcgaacatctccgacct	242
QY	279	ggttggaaacgagactgtcaatgtgggagacagcccaaaccttgatgttctacagaatagatccg	338
Db	243	ggttggaaacgagactgtcaatgtgggagacagcccaaaccttgatgttctacagaatagatccg	302
QY	339	cttctgtcccaacgagcgttcttcatttgaggaacattcttaagaactgtgacgagacatata	398
Db	303	cttctgtcccaacgagcgttcttcatttgaggaacattcttaagaactgtgacgagacatata	362
QY	399	ctctcttgaagacaacatcactccatcacacagtgagctgttctcctctgtcgagaaacagagat	458
Db	363	ctctcttgaagacaacatcactccatcacacagtgagctgttctcctctgtcgagaaacagagat	422
QY	459	gaactgtgactgtccaaacccctcactcgtcagagaggtgtgatttaaaagctcccaagga	518
Db	423	gaactgtgactgtccaaacccctcactcgtcagagaggtgtgatttaaaagctcccaagga	482
QY	519	gat-ccgaagagcggc-tgtccgggacctcagagatcgaagcgtgggcttctaaggatccgg	576
Db	483	gatcccaagagcggcgtattgtccgggacctcagagatcgaagcgtgggcttctaaggatccgg	542
QY	577	ctggaagagcagagcagcagcgga-cggtgtggcgcagacagagaactccagaagcgtttcca	635
Db	543	ctggaagagcagagcagcagcgga-cggtgtggcgcagacagagaactccagaagcgtttcca	602
QY	636	gaa--cctgaactgtgcgacgacacaaacactccgcatcacaacgacatcccaagtgcgg	693
Db	603	gaa-cttgaactgtgcgacgacacaaacactccgcatcacaacgacatcccaagtgcgg	662
QY	694	tgtagcgtgagcgtcagaca	713
Db	663	gtgaccttgaggccttcagaca	682
RESULT	5		
LOCUS	AM338856/c		
DEFINITION	ha67f02.x1 NCI-CGAP_Panl Homo sapiens cDNA clone IMAGE:2878779 3'		
	similar to TR:096029 096029 7-60.; contains element MSRI repetitive		
	element ;, mRNA sequence.		
ACCESSION	AM338856		
VERSION	AM338856.1		
KEYWORDS	GI:6835482		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 606)		
AUTHORS	NCI-CGAP		
TITLE	http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Life Technologies catalog #: 11548-013		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/ILMW at:		
	www-bio.ljll.gov/bdnp/image/image.html		
	Seq primer: -40UP from Glibco		
	High quality sequence stop: 421.		
FEATURES	Location/Qualifiers		
SOURCE	1..606		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2878779"		
	/clone_lib="NCI-CGAP_Panl"		
	/tissue_type="adenocarcinoma"		

[illegible]

|||||
Db 421 GCCGAGAGCCCATCGAGAGCCCATGCCAGCCCGCGAGGACGTACAGAGGATGAGCA 480
Oy 1656 gccgagagcccatcgagagaccagagcccgcccgagagagcccg 1741
|||||
Db 481 GCCGAGAGCCCATCGAGAGCCCATGCCAGCCCGCGAGGACGTG 526
|||||
RESULT 8
LOCUS BE410591 591 bp mRNA EST 21-JUL-2000
DEFINITION 60130331JF1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637634 5',
mRNA sequence.
ACCESSION BE410591
VERSION BE410591.1 GI:9347041
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 591)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM337 row: f column: 03
High quality sequence start: 591.
High quality sequence stop: 591.
Location/Qualifiers
1. 591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3637634"
/clone_1lb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 136 a 149 c 196 g 109 t 1 others
ORIGIN
Query Match 21.0%; Score 506.2; DB 35; Length 591;
Best Local Similarity 99.4%; Pred. No. 1.1e-92;
Matches 508; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

|||||
Db 261 GCCAGAGGAGCAATGTAGATGAGGACCAACTATCCGATTCGTGTAACGAGACTGC 320
Oy 295 aatggagacagcccaactgaattctacagaatgatgatcgcttcctcccaagc 354
|||||
Db 321 AATGGGACAGCCCAACTGATTTTACAGAAATGATGATCCCTTCTCCCAACGGC 380
|||||
Oy 355 ttttcattgaggaattcttcagactgagcaggaactatgacctcttgaggaat 414
|||||
Db 381 TGTTCATTGAGGACATCTTCAGACTGAGAGCAACTATGACTCTTGAGGACAT 440
|||||
Oy 415 cactctacacacagctgctgttctctctcgagagaccagagatgagctgcatgcaag 474
|||||
Db 441 CACTCTCATTCAGACTGCTCTTCTCTGAGAGAACGAGAGTGCATGCCAGCCAG 500
|||||
Oy 475 cccctacagctcaggaaggtcgaggtgtttaaaagctccagagatccagagcgctt 534
|||||
Db 501 CCCCTCAGCGCTCAGGAGAGTGGAGGTGTTAAAGCTCCAGAGATCCAGAGCGGCTT 560
|||||
Oy 535 gtcgggacctcagagctatgctgcttc 565
|||||
Db 561 GTCCGGGCTCAGGACTATCTGCTGGCTTCT 591
|||||
RESULT 9
LOCUS AM246100 546 bp mRNA EST 07-JAN-2000
DEFINITION 2821312.5prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821312 5',
mRNA sequence.
ACCESSION AM246100
VERSION AM246100.1 GI:6589093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other-ESTs: 2821312.3prime
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTF CDNA Library Preparation: Ling
Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LNL) DNA Sequencing by: Berkeley MGC sequencing
project clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/brp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PHRAP suite. Poly-T identification: patmatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:
http://www.genome.washington.edu
Plate: LNCM6 row: H column: 17
High quality sequence stop: 470.
Location/Qualifiers
1. 546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2821312"
/clone_1lb="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit

BASE COUNT 122 a 210 c 184 g 30 t
ORIGIN (stratagene) and Superscript II RT (Life Technologies)."

Query Match 20.7%; Score 498; DB 21; Length 546;
Best Local Similarity 95.4%; Pred. No. 4.8e-91;
Matches 513; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1459 gcccaagccttgcccttcggtccctccatcggggaccccaagctggacac 1518
D 9 GCCCAACACTTGGCCCTTCCGGTCCCTGCCCATCGGGGCAACCCCAAGCTGGACAC 68
QY 1519 aaggaagaagggttgaggagacagaaggtcgaaggggcccagaagtaccct 1578
D 69 AGTGAACACGGGTTGAGGAGACAGAAAGTGCAGAGGGGCCCAAGAGGTACCCCT 128
QY 1579 gggagccatcggagagcccaaggcccaagcagacactcaggggagcagcc 1638
D 129 GGGAGCCATCGAGACCCAGGCCCAAGCCACAGACTTCAGGGGACGACCCAGCC 188
QY 1639 aagaaccatcggagagcccaaggcccaagcagcactacaaggatgaagcagcc 1698
D 189 GAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGACGAGCCAGCC 248
QY 1699 gagagccatcggagagcccaaggcccaagcagcactcgaaggagcagccagcc 1758
D 249 GAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGACGAGCCAGCC 308
QY 1759 gagagccatcggagagcccaaggcccaagcagcactcgaaggagcagccagcc 1818
D 309 AAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGATGAGCCAGCC 368
QY 1819 aagatccatcggagagcccaaggcccaagcagcactacaaggatgaagcagcc 1878
D 369 GAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGACGAGCCAGCC 428
QY 1879 gagagccatcggagagcccaaggcccaagcagcactcgaaggagcagccagcc 1938
D 429 GAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGACGAGCCAGCC 488
QY 1939 gagagccatcggagagcccaaggcccaagcagcactcgaaggagcagccag 1996
D 489 GAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGATGAGCCAGCC 546

RESULT 10
BE207411/c 551 bp mRNA EST 27-JUN-2000
LOCUS ba66d10.x1 NIH_MGC.20 Homo sapiens cDNA clone IMAGE:2905363 3'
DEFINITION similar to TR:096029 096029 7-60. ; mRNA sequence.
ACCESSION BE207411
VERSION BE207411.1 GI:8750809
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 551)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMU at:
www-bio.11nl.gov/dbip/image/image.html
Seq primer: -40UP from Glibco

FEATURES High quality sequence stop: 434.
Location/Qualifiers
1..551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2905363"
/clone_1id="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site: 1; XhoI; Site: 2;
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EORI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 96 a 175 c 180 g 100 t
ORIGIN

Query Match 19.8%; Score 477.4; DB 34; Length 551;
Best Local Similarity 99.8%; Pred. No. 7.1e-87;
Matches 478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1930 gagcagcccgagagccatcggagagcccaagcccccgcgagagcctcagggagc 1989
D 551 GAGCAGCCGAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGAT 492
QY 1990 gagcagcccgagagccatcggagagcccaagcccccgcgagagcctacaaggat 2049
D 491 GAGCAGCCGAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGAT 432
QY 2050 gagcagcccgagagccatcggagagcccaagcccccgcgagagcctcagggagc 2109
D 431 GAGCAGCCGAGAGCCATCGAGACCCAGGCCCCCGCCAGCAGACTTCAGGGGAT 372
QY 2110 aagctcggagagccttaaggagagagtgccgtcggtcgttcgttcctcgtccgc 2169
D 371 AAGCTGGAGAGCCCTTAAGAAAGAGTGCCTGCGCGCTCTGCTCTGCTCTGCTGC 312
QY 2170 tgcagggcctggggcccccagagctgtcgagggtcccccagagccttgcttgagcc 2229
D 311 TGCAAGGGCTGGGGCTCCGAGCTGCTGGGGCTCCCTCAAGCTTGTGTGACCC 232
QY 2230 gtgaccatgaccacagctgctgctcctgtgtgggacatataagcaccagagcc 2289
D 251 GTGACCATGACCCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 192
QY 2290 gcgagggccttaaggagagcccaagcctcgaaggagcctcgtgctgtctcc 2349
D 191 GCGAGGCCCTCAAGGAGCCCAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 132
QY 2350 caccagcctcctccctgagccctgtcttgaattgacccctcgaaggaggggagcg 2408
D 131 CACCAGCTCTCCCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 73

RESULT 11
BE336852 493 bp mRNA EST 14-JUL-2000
LOCUS ba97612.y1 NIH_MGC.14 Homo sapiens cDNA clone IMAGE:2957902 5'
DEFINITION similar to TR:096029 096029 7-60. ; mRNA sequence.
ACCESSION BE336852
VERSION BE336852.1 GI:9189237
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 493)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)


```

/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      327 a 416 c 389 g 164 t
ORIGIN
```

```

Query Match      18.2% Score 438.8; DB 35; Length 1296;
Best Local Similarity 84.3%; Pred. No. 4,8e-79;
Matches 666; Conservative 0; Mismatches 92; Indels 32; Gaps 14;

QY 515 aggaagatccagagac-ggctgtccgggacctacagactcagctcagctcagcagc 573
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 11 AGGAGATCCAGAGAGTAGCTGTGCGGGCTACGACTCATGCTGGCTTCTACGGGATC 70

QY 574 cggctgagagaccgagcagc-gcacggtggcgagcagcagcagcagcagcagc 632
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 71 CGGCTGAGAGACCGAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 130

QY 633 cgaagaccctgactgagcagcagcagcagcagcagcagcagcagcagcagcagc 692
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 131 CCAGAACCTGAACTGGCGGAGCCACACACACCTCCGATCACAGCATCTCAAGTCGCT 190

QY 693 gtgtgagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 752
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 191 GGGTGAAGCTGGGCTGAGACACTTCCAGGCGCGCTGCTTCTTCTTCCAGAGAGAC 250

QY 753 gctgtgctggcgagcagcagcagcagcagcagcagcagcagcagcagcagc 811
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 251 GGTGCTGGCGGCGGAGCTCCGGGGGTGGCGGAGAGTCCCTGACTTCTTCTTCTTCA 310

QY 812 cgtgtgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 870
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 311 CGGTGGCTGGCGGAGCTGAGACACTTCCAGGCGCGCTGCTTCTTCTTCCAGAGAG 370

QY 871 cccgcg-tgcaagctcgt--ctggggggcccccaagcagcagcagcagcagcagc 927
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 371 ACCCGATGCAATGCTGTAATGGGGCCCCACAGAGAGAGTCCGAGAGTTCAGAGCC 430

QY 928 tctctgcagcat-cgcgtcagaggtcctcagaggtgagaggtgagaggtgagag 986
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 431 TCTCTGCCCATACCGCTCGAGGGCTCCAGGAAGTGGAGAGAGAGAGAGAGAGCC 490

QY 987 cccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1040
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 491 CCCCGACCAAGAGGCGAGACCCAGGCTGCGACCTGTGGGCCAGAGCCATAGCAAG 550

QY 1041 ggggtggggcagaggtgagagaggtggcccgagcagcagcagcagcagcagc 1097
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 551 GCGGGCAGCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 610

QY 1098 gggagccctggagagagcagcagcagcagcagcagcagcagcagcagcagc 1154
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 611 GGAGCAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 670

QY 1155 ggaag-----ccttaagccccaagagagcagagagagagagagagagag 1204
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 671 GAGAGAGCCCATATAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 730

QY 1205 ggggagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1264
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 731 CGGAGAGAGATCCG-CCACAGAGCCAGGCGCTCAGAGTGCCTCA-AGGTGCAAGAG 788

QY 1265 cctcgaattt 1274
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 789 CACTGAACCT 798
```

```

RESULT 15
BE391750
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
http://image.llnl.gov
Plate: LHC252 row: f column: 03
High quality sequence start: 110
High quality sequence stop: 686.
location/Qualifiers
1..757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3604994"
/clone_1b="NIH-MGC 44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      170 a 204 c 251 g 131 t 1 others
ORIGIN

Query Match      18.1% Score 435.4; DB 35; Length 757;
Best Local Similarity 92.9%; Pred. No. 2.3e-78;
Matches 577; Conservative 0; Mismatches 31; Indels 13; Gaps 11;

QY 51 cgaactcactctggggagagcagcagcagcagcagcagcagcagcagcagcagc 110
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 94 CAATTCACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 153

QY 111 ggaagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 170
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 154 GAGAGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 213

QY 171 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 229
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 214 GAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 273

QY 230 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 287
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 274 GGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333

QY 288 agactgcaattgggagacagcagcagcagcagcagcagcagcagcagcagcagc 345
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 334 AGACTGCAATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
```

```

QY 346 cccaa-cggctgttcatttgaaggaatcttcagaactggacggacaactatgacctcct 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 CCCAAGCGGCTGTTCATTGAGACATTTCTCAGAACTGGAGGACACTATGACTCTCT 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 405 tg-aggaacaatcaactcctacatccagltggtgttccctctgagaaacaggaatgaa 463
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 TGCAGGACAAATCACTCTACATCCAGTGGTGCTCTCCTCTGCGAGACCAAGAGTGA 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 464 ggcattgccaagcccc-tcagctcaggga-gtctgagggtgtttaaagctccagagat 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 GGCATGCCAAAGCCCTTCACGCTCAGGAGCGTCTAGAGGCTCCAGAGAT 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 522 ccaaggagcggctgtccgggacctagaagctcat-gctgggcttctacgggatcc-ggctg 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 CCAGGAGCGGCTTCTCCGGGCTTACGAGCTTCATAGCTGGCTTCTACGGGATCCGGGCTG 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 580 gaagaccgaagcagc--ggcacggtggccgagacagaactacagaagcgtctccaga 637
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 GAGGACCGAGGACGCGCCAGCGGTGGCCGAGCAGAACTACAGAGGCGCTTCCAGA 693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 acctgaactggcgagccaca 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 694 GCTTAAGTGGGACGACCAAA 714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: October 12, 2000, 12:02:37
 Job time: 24145 sec

XX 20-FEB-1992.
XX
XX 01-AUG-1991; 91WO-US05463.
XX
XX 02-AUG-1990; 90US-0561968.
XX
XX (COLS) UNIV OF COLORADO.
XX
XX Gold L, Tuerk C;
XX
XX WPI; 1992-080018/10.
XX
XX New method of systematic polypeptide evolution by reverse
XX translation - by linking each polypeptide in sample mixt. to
XX individualised mRNA allowing further synthesis of selected
XX polypeptide(s)
XX
XX Example; Page 55; 102pp; English.
XX
XX The sequence is that of an example randomising oligonucleotide which
XX is used in the prepn. of mRNA encoding candidate polypeptides for the
XX method of systematic polypeptide evolution by reverse translation
XX (SPERT). The method provides a rapid way of isolating and identifying
XX polypeptide ligands which bind to target mols. The polypeptide ligands
XX can be used in e.g. assay methods, diagnostic procedures, cell sorting,
XX as activators or inhibitors of target mol. function, as probes, as
XX sequestering agents, drug delivery vehicles, modifiers of hormone
XX action and as catalysts. See also Q21830-Q21832.
XX
XX Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

Query Match 3.9%; Score 62.6; DB 13; Length 390;
Best Local Similarity 65.2%; Pred. No. 0.0019;
Matches 92; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 33 catgagacagccagctcgactccactggagagagagagagatgctggagagc 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 catgagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 64
QY 93 ggaagacagagactcgagagcagcgccgagcgcgagcgagcgagcgagcgagc 152
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 cgacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 124
QY 153 cgagagcagagagtcgagcagcagcagcagcagcagcagcagcagcagcagc 173
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 cgacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 145

RESULT 4
Q36859
ID Q36859 standard; DNA; 390 BP.
XX
XX Q36859;
XX
XX 22-JUN-1993 (first entry)
XX
XX PCR primer for 5' fixed sequence contg. T7 promoter and RBS.
XX
XX Systematic peptide evolution by reverse translation; SPERT; ligand;
XX specific; inhibitors; probes; assay; cell sorting; ss.
XX
XX Synthetic.
XX
XX WO9303172-A.
XX
XX 18-FEB-1993.
XX
XX 31-JAN-1992; 92WO-US00801.
XX
XX 01-AUG-1991; 91US-0739055.
XX

PA (UYRE-) UNIV RES CORP.
XX
XX Gold L, Pribnow D, Smith JD, Tuerk C;
XX
XX WPI; 1993-076529/09.
XX
XX Systematic polypeptide evolution by reverse translation - used
XX for prodn. of polypeptide ligand specific for desired target
XX molecule
XX
XX Example 1; Page 84; 98pp; English.
XX
XX SPERT is used to select novel polypeptides that bind the antibody
XX of the epitope commonly recognised by the antisera from autoimmune
XX mice which are the F1 progeny of a cross of NZB and NZW parents
XX (Portanova et al., J. Immunol. 144, 4633, 1990). The known epitope
XX consists of ca. 10 amino acids at the N-terminus of the histone H2B
XX protein. To make mRNA encoding candidate polypeptides a 5' fixed
XX sequence composed of a T7 promoter sequence and a ribosome binding
XX site which is recognised by both prokaryotic and eukaryotic ribosomes,
XX terminating in a restriction endonuclease site is synthesised and cloned
XX using a number of oligonucleotides (example shown). A 3' fixed sequence
XX is placed into a restriction site to provide an mRNA encoding the C-
XX terminal trailer sequence of ca. 100 nucleotides lacking stop codons.
XX In addition, a 3' primer annealing site is provided so that cDNA
XX synthesis can be accomplished on the mRNA recovered from partitioned
XX ribosome complexes. See also Q36845-63.
XX
XX Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

Query Match 3.9%; Score 62.6; DB 14; Length 390;
Best Local Similarity 65.2%; Pred. No. 0.0019;
Matches 92; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 33 catgagacagccagctcgactccactggagagagagagagatgctggagagc 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 catgagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 64
QY 93 ggaagacagagactcgagagcagcgccgagcgcgagcgagcgagcgagcgagc 152
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 cgacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 124
QY 153 cgagagcagagagtcgagcagcagcagcagcagcagcagcagcagcagcagc 173
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 cgacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 145

RESULT 5
Q73500
ID Q73500 standard; DNA; 8438 BP.
XX
XX Q73500;
XX
XX 15-MAY-1995 (first entry)
XX
XX DNA encoding Pseudorabies virus large latency transcript.
XX
XX Pseudorabies virus; PRV; LTV; large latency transcript;
XX attenuated virus; vaccine; early protein O; EPO; HSV-1 ICPO;
XX protecting animals; deletion mutants; swine; ds.
XX
XX Pseudorabies virus.
XX
XX OS
XX Key Location/Qualifiers
XX FH 1..7013
XX FT misc_feature
XX FT /*tag- a
XX FT /*note- "derived from PRV strain Infn"
XX FT 7014..8425
XX FT /*tag- b
XX FT /*note- "derived from PRV strain Ka"
XX FT 622..6498
XX FT CDS
XX FT /*tag- c


```

ID V19941 standard; DNA: 137507 BP.
XX
AC V19941.
XX
DT 03-AUG-1998 (first entry)
XX
DE KSHV long unique coding region and terminal repeat.
XX
KW KSHV: HHV8: human herpes virus 8; macrophage inflammatory protein II;
interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
complement-binding protein; glycoprotein; capsid protein IV; infection;
lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
HIV immune status; anti-inflammatory agent; therapy; ds.
XX
OS Kaposi's sarcoma-associated herpes virus.
XX
FH Key
FH Location/Qualifiers
FT 1142..2794
FT /tag= a
FT /product= complement-binding protein
FT 8699..11236
FT /tag= b
FT /product= glycoprotein B
FT complement (17261..17875)
FT /tag= c
FT /product= interleukin 6
FT complement (21548..21832)
FT /tag= d
FT /product= macrophage inflammatory protein II
FT complement (27137..27424)
FT /tag= e
FT /product= interferon regulatory factor 1
FT 2861..29741
FT /tag= f
FT /product= protein TI.1
FT complement (58976..60175)
FT /tag= g
FT /product= glycoprotein M
FT complement (69412..69915)
FT /tag= h
FT /product= glycoprotein L
FT complement (88410..88910)
FT /tag= i
FT /product= interferon regulatory factor 2
FT 89600..90541
FT /tag= j
FT /product= interferon regulatory factor 3
FT 90173..90643
FT /tag= k
FT /product= glycoprotein X
FT complement (93636..94127)
FT /tag= l
FT /product= interferon regulatory factor 4
FT complement (111931..112443)
FT /tag= m
FT /product= capsid protein IV
FT complement (123808..127296)
FT /tag= n
FT /product= immediate early protein
XX
PN M09804576-A1.
XX
PD 05-FEB-1998.
XX
PE 22-JUL-1997; 97MO-US13346.
XX
PR 29-NOV-1996; 96US-0757669.
PR 25-JUL-1996; 96US-0686243.
PR 25-JUL-1996; 96US-0686348.
PR 25-JUL-1996; 96US-0686350.
PR 25-JUL-1996; 96US-0687253.
PR 25-JUL-1996; 96US-0688814.

```

```

PR 05-SEP-1996; 96US-0708678.
PR 10-OCT-1996; 96US-0728323.
PR 13-NOV-1996; 96US-0747887.
PR 13-NOV-1996; 96US-0748640.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JT;
XX
DR WPI: 1998-130615/12.
XX
PT New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients
XX
PS Example 2; Page 135-203; 230pp; English.
XX
CC This sequence represents the long unique region and terminal repeat of
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC invention which encode KSHV polypeptides selected from: (a) viral
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC by it, and antibodies (Ab) specific for the proteins are useful for
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC fluids or tissue samples. HHV8 infections can be treated with antisense
CC or triplex forming molecules or agents that bind specifically to the
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC while the protein can be used in protective vaccines. Ab may also be used
CC to differentiate between lymphomas, and HHV8 may be implicated in many
CC other lymphoproliferative diseases such as lymphomas, Leukemia,
CC splenomegaly and mycosis fungoides. Cells and animals containing the
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC can be inhibited with methotrexate. These can also be used to determine
CC the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP II may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC 81 open reading frames.
XX
SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

Query Match 3.8%; Score 60.2; DB 19; Length 137507;
Best Local Similarity 57.8%; Pred. No. 0.0082;
Matches 107; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 2 agaattcagcgccgctgaattcagccgagcagcagcccgactgcactccacct 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126299 AGGATGACGAGGAGGATGACGAGGATGACGAGGATGACGAGGATGACGAGG 126210

QY 62 gggaggagagcagagatcgaggagcagcgagagcagagactgcagagcgagag 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126209 AGGATGACGAGGAGGATGACGAGGATGACGAGGATGACGAGGATGACGAGG 126150

QY 122 ccgcccgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126149 AGGATGACGAGGAGGATGACGAGGATGACGAGGATGACGAGGATGACGAGG 126090

QY 182 cggcg 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126089 AGGAG 126085

RESULT 8
V62176/C
ID V62176 standard; DNA: 117213 BP.
XX
AC V62176;
XX
DT 13-JAN-1999 (first entry)

```

```
XX DE HSV-2 strain SBS Contlg ID 15 DNA sequence.
XX KM HSV-2 strain SBS; immunological response induction; therapy;
XX KW antiviral identification; viral protein inhibitor; ss.
XX OS Herpes simplex virus type 2.
FH Location/Qualifiers
FT CDS 755..1297
FT     /tag= a
FT     /product= "ORF#1 protein"
FT     /note= "encoded protein shown in W72170"
FT     1170..2174
FT     /tag= b
FT     /product= "ORF#2 protein"
FT     /note= "encoded protein shown in W72171"
FT     2229..2930
FT     /tag= c
FT     /product= "ORF#3 protein"
FT     /note= "encoded protein shown in W72172"
FT     complement (3130..3735)
FT     /tag= d
FT     /product= "ORF#4 protein"
FT     /note= "encoded protein shown in W72173"
FT     complement (3802..6447)
FT     /tag= e
FT     /product= "ORF#5 protein"
FT     /note= "encoded protein shown in W72174"
FT     6017..8482
FT     /tag= f
FT     /product= "ORF#6g protein"
FT     /note= "encoded protein shown in W72250"
FT     6026..8482
FT     /tag= g
FT     /product= "ORF#6f protein"
FT     /note= "encoded protein shown in W72249"
FT     6065..8482
FT     /tag= h
FT     /product= "ORF#6e protein"
FT     /note= "encoded protein shown in W72248"
FT     6167..8482
FT     /tag= i
FT     /product= "ORF#6d protein"
FT     /note= "encoded protein shown in W72247"
FT     6296..8482
FT     /tag= j
FT     /product= "ORF#6c protein"
FT     /note= "encoded protein shown in W72246"
FT     6326..8482
FT     /tag= k
FT     /product= "ORF#6b protein"
FT     /note= "encoded protein shown in W72245"
FT     6446..8482
FT     /tag= l
FT     /product= "ORF#6a protein"
FT     /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT     /transl_except= (pos: 7481..7486, aa: Ile)
FT     /note= "encoded protein shown in W72244"
FT     8457..9347
FT     /tag= m
FT     /product= "ORF#7 protein"
FT     /note= "encoded protein shown in W72175"
FT     complement (9604..11855)
FT     /tag= n
FT     /product= "ORF#8 protein"
FT     /note= "encoded protein shown in W72176"
FT     complement (11905..14508)
FT     /tag= o
FT     /product= "ORF#9b protein"
FT     /note= "encoded protein shown in W72222"
FT     complement (11905..14520)
FT     CDS
FT     /tag= p
FT     /product= "ORF#9a protein"
FT     /note= "encoded protein shown in W72223"
FT     14399..15802
FT     /tag= q
FT     /product= "ORF#10 protein"
FT     /note= "encoded protein shown in W72177"
FT     complement (15996..16286)
FT     /tag= r
FT     /product= "ORF#11 protein"
FT     /note= "encoded protein shown in W72178"
FT     complement (16202..18064)
FT     /tag= s
FT     /product= "ORF#12 protein"
FT     /note= "encoded protein shown in W72179"
FT     complement (18105..19661)
FT     /tag= t
FT     /product= "ORF#13 protein"
FT     /note= "encoded protein shown in W72180"
FT     complement (19415..20074)
FT     /tag= u
FT     /product= "ORF#14 protein"
FT     /note= "encoded protein shown in W72181"
FT     20155..21453
FT     /tag= v
FT     /product= "ORF#15 protein"
FT     /note= "encoded protein shown in W72182"
FT     complement (21326..22291)
FT     /tag= w
FT     /product= "ORF#16 protein"
FT     /note= "encoded protein shown in W72183"
FT     complement (22546..24654)
FT     /tag= x
FT     /product= "ORF#17 protein"
FT     /note= "encoded protein shown in W72184"
FT     24684..25955
FT     /tag= y
FT     /product= "ORF#18 protein"
FT     /note= "encoded protein shown in W72185"
FT     complement (26295..27251)
FT     /tag= z
FT     /product= "ORF#19 protein"
FT     /note= "encoded protein shown in W72186"
FT     complement (27630..31754)
FT     /tag= aa
FT     /product= "ORF#20b protein"
FT     /note= "encoded protein shown in W72225"
FT     complement (27630..31784)
FT     /tag= ab
FT     /product= "ORF#20a protein"
FT     /note= "encoded protein shown in W72224"
FT     complement (32067..32735)
FT     /tag= ac
FT     /product= "ORF#21 protein"
FT     /note= "encoded protein shown in W72187"
FT     33140..34984
FT     /tag= ad
FT     /product= "ORF#22a protein"
FT     /note= "encoded protein shown in W72226"
FT     33386..34984
FT     /tag= ae
FT     /product= "ORF#22b protein"
FT     /note= "encoded protein shown in W72227"
FT     complement (35205..37721)
FT     /tag= af
FT     /product= "ORF#23 protein"
FT     /note= "encoded protein shown in W72188"
FT     complement (38058..39188)
FT     /tag= ag
FT     /product= "ORF#24 protein"
FT     /note= "encoded protein shown in W72189"
FT     39090..39935
FT     /tag= ah
```

FT	/product= "ORF#25 protein"
FT	/note= "encoded protein shown in W72190"
FT	40216..41973
FT	/tag= al
FT	/product= "ORF#26 protein"
FT	/note= "encoded protein shown in W72191"
FT	42206..44178
FT	/tag= aj
FT	/transl_except= (pos: 44063..44064, aa: Lys)
FT	/product= "ORF#27 protein"
FT	/note= "encoded protein shown in W72192"
FT	complement (44853..47297)
FT	/tag= ak
FT	/product= "ORF#28 protein"
FT	/note= "encoded protein shown in W72193"
FT	47122..47338
FT	/tag= al
FT	/product= "ORF#29 protein"
FT	/note= "encoded protein shown in W72194"
FT	complement (47305..49662)
FT	/tag= am
FT	/product= "ORF#30 protein"
FT	/note= "encoded protein shown in W72195"
FT	complement (50035..51666)
FT	/tag= an
FT	/product= "ORF#31 protein"
FT	/note= "encoded protein shown in W72196"
FT	complement (51701..53575)
FT	/tag= ao
FT	/product= "ORF#32 protein"
FT	/note= "encoded protein shown in W72197"
FT	54393..58115
FT	/tag= ap
FT	/product= "ORF#33 protein"
FT	/note= "encoded protein shown in W72198"
FT	complement (58060..58977)
FT	/tag= ag
FT	/product= "ORF#34 protein"
FT	/note= "encoded protein shown in W72199"
FT	complement (58970..60760)
FT	/tag= ar
FT	/product= "ORF#35 protein"
FT	/note= "encoded protein shown in W72200"
FT	60759..61151
FT	/tag= as
FT	/product= "ORF#36 protein"
FT	/note= "encoded protein shown in W72201"
FT	61241..62071
FT	/tag= at
FT	CDS
Query Match	3.7%; Score 60; DB 19; Length 117213;
Best Local Similarity	67.7%; Pred. No. 0.0088;
Matches	84; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
OY	63 gggaggaaggcgcgagatgtcggaagacgcggaagacgactctcgaggaacgcgcgagtc 122 Db 113738 GGAGGAGGAGCGCGGAGGAGGAGGAGCGCGAGAAGCAGCGAGAGAGAGAGGAGCGC 113739
OY	123 cgcgcgagcgaggaagcgagcaagcgaggaagcaggaagcagtggtcggaagaacgcgcgagtc 182 Db 113738 GGAGGAGGAGCGCGGAGGAGGAGCGCGAGAAGCAGGAGCGGAGGCGCGGGAGGA 113679
OY	183 ggccg 186 Db 113678 GGAG 113675
RESULT	9
ID	V64956 standard; DNA: 150 BP.
XX	V64956
AC	V64956;
XX	

DT	05-FEB-1999	(first entry)
XX		
DE	Mouse histone H2B protein epitope oligonucleotide fragment #7.	
XX		
XX	Ligand; assay; diagnostic; inhibitor; probe; cell sorting; catalysis;	
KW	sequestering agent; epitope; ss.	
XX		
OS	Synthetic.	
OS	Mus sp.	
XX		
PN	US5843701-A.	
XX		
PD	01-DEC-1998.	
XX		
PF	31-JAN-1992; 92US-0829461.	
XX		
PR	31-JAN-1992; 92US-0829461.	
PR	02-AUG-1990; 90US-0561968.	
PR	01-AUG-1991; 91US-0739055.	
XX		
PA	(NEXS-) NEXSTAR PHARM INC.	
XX		
PI	Gold L, Pridnow D, Smith JD, Tuerk C;	
XX		
DR	WPI; 1999-044575/04.	
XX		
PT	Production of polypeptide ligands using mRNA-polypeptide conjugates	
PT	- by affinity selection, RNA amplification and polypeptide synthesis	
XX		
PS	Example 10; Column 45-46; 35pp; English.	
XX		
CC	V64950-V64972 are sequences used in a method for the production of a	
CC	polypeptide ligand of a target molecule. The polypeptide products can	
CC	be used in assays, in diagnostic procedures, in cell sorting, as	
CC	inhibitors of target molecule function, as probes, as sequestering	
CC	agents, or may have catalytic activity.	
XX		
SO	Sequence 150 BP; 44 A; 46 C; 53 G; 6 T; 1 other;	

[illegible]

```
XX 10-JUL-1997.
PD 30-DEC-1996; 96WO-US20785.
XX 12-APR-1996; 96US-0632175.
XX 29-DEC-1995; 95US-0009409.
PR 29-DEC-1995; 95US-0580539.
PR 30-JAN-1996; 96US-0010835.
PR 30-JAN-1996; 96US-0594242.
XX (DARM-) DARWIN MOLECULAR CORP.
PA (OSHI-) OSHIMA J.
PI Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;
XX WPI; 1997-363671/33.
DR Isolated nucleic acid molecule encoding the WRN gene product -
XX useful for detection and treatment of Werner's syndrome, and related
PT diseases
PT Claim 1; Fig 7; 153pp; English.
XX This sequence represents a fragment of the genomic sequence containing
CC the coding region for the mouse WRN gene (X83004). The corresponding
CC human gene (X83001) encodes a protein related to Werner's syndrome.
CC The products can be used for the detection and treatment of Werner's
CC syndrome (WS), an autosomal recessive disorder with a complex phenotype,
CC as well as related diseases.
XX Sequence 51259 BP; 14533 A; 9635 C; 10266 G; 16825 T; 0 other;
SQ
Query Match 3.6%; Score 58.4; DB 18; Length 51259;
Best Local Similarity 63.6%; Pred. No. 0.016;
Matches 89; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 47 actggactccacccctgtagagagagagagatgtagcggagagacggagacgact 106
DB 279 ACAACAATACATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 220
QY 107 ggcagagagcgagagcgccgagcgagagagcgagagcgagagagagagagagat 166
DB 219 AGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 160
QY 167 cggagagagcgagcgagcgagcg 186
DB 159 AGGAGAGCAGAGAGAGAGAG 140
RESULT 11
V55831
ID V55831 standard; DNA; 799 BP.
AC V55831;
XX 18-NOV-1998 (first entry)
DT Nucleotide sequence of the stabilizing sequence-encoding insert.
DE
XX Fusion protein; stabilizing polypeptide; proteolytic degradation;
XX resistance; half-life; autoimmune disease; inflammation; nitro drug;
XX IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
XX nitroreductase protein; enzyme therapy; produg therapy; protease;
XX cancer; pathological condition; ss.
OS Epstein-Barr virus.
XX MO9822577-A1.
XX 28-MAY-1998.
```

```
PF 17-NOV-1997; 97WO-IB01508.
XX 25-JUN-1997; 97US-0048945.
PR 15-NOV-1996; 96US-0030986.
XX (MASU/) MASUCCI M G.
XX Masucci MG;
XX WPI; 1998-312463/27.
DR New fusion proteins resistant to proteolytic degradation -
XX comprising a core protein with a stabilizing polypeptide comprising
PT a peptide sequence containing glycine repeats
PT Disclosure; Fig 4B; 120pp; English.
XX This is a nucleotide sequence of the stabilizing sequence-encoding
CC insert. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilizing polypeptide of
CC formula (Glya)(Glyb)(Glyc)2ln where Glya, Glyb, Glyc are 1-6
CC sequential gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilizing polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
XX Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;
SQ
Query Match 3.6%; Score 57.2; DB 19; Length 799;
Best Local Similarity 43.2%; Pred. No. 0.02;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;
QY 718 caagcgccactggtccgcttctctctgagagagcgctgtcggcgagagctccggg 777
DB 66 caggagcagagagagagagagcagagcagagagagagagagagagagagagagag 125
QY 778 gtgcgagagagtccttgactactcatgtccgctgcgctgcgcagaccagcgccg 837
DB 126 gggcagagagcagagagagagagagcagagagagagagagagagagagagagag 185
QY 838 cagctgtgctacttcgctcgtgagcattccgcccgcctgcgaattcgtctgtggggcc 897
DB 186 gagcagagagagagagagagagagagagagagagagagagagagagagagagag 245
QY 898 caagacaagctcgcgagaggttcaagccagctcttcgcccgcattcgtctgagggctcag 957
DB 246 caggagagagcagagagagagagagagagagagagagagagagagagagagag 305
QY 958 aaggtgagagagagagagagagagagagagagagagagagagagagagagagag 1017
DB 306 gggcagagagagagagagagagagagagagagagagagagagagagagagagag 365
QY 1018 acctgtggccagagcagatagcaaggtgtggggcagaggtgcagagagagagagag 1077
DB 366 gagcagagagagagagagagagagagagagagagagagagagagagagagagag 425
QY 1078 agcgtggagagagagagagagagagagagagagagagagagagagagagagagag 1137
DB 426 gggcagagagagagagagagagagagagagagagagagagagagagagagagag 485
QY 1138 cagcgagagagagagagagagagagagagagagagagagagagagagagagagag 1197
```


Dh 486 gagggcagagcagagagggcagagagggcagagagggcagagagggcagag 545
Cc 1198 ctgagccggcgagggagccgcccacagagccctcctcagtgctcagagtgag 1257
Dh 546 caggaagggcgagggcagagcagagagggcagagggcagagagcagagagag 605
Qy 1258 aagatcctcctgaatttgaggggtgtgcccctcagccagagccctcagagc 1317
Dh 606 gggcagagagggcgagagcagagagggcgagagggcgagagggcgagagaggg 665
Qy 1318 caggaagtggcggtcagagaccctggggag 1347
Dh 666 caggaagcagagggcgagagagggcagagag 695

RESULT 12
X90923/c
ID X90923 standard; DNA; 5452 BP.
XX
XX X90923;
AC
DT 17-JAN-2000 (first entry)
XX
DE Anti-sense strand of pcwvEBNA plasmid.
XX
KW Anti-sense strand: plasmid pcwvEBNA; EBNA 1; episome; transfection;
KW Epstein Barr Virus Nuclear Antigen 1; origin of replication;
KW EBV oriP; eucaryotic host cell; recombinant cell line; ion channel;
KW multiple gene expression; receptor; transporter protein; gene therapy;
KW transcription factor; adhesion molecule; antisense therapy;
KW gene amplification; cell immortalisation; ss.
XX
OS Epstein-Barr virus.
OS Cytomegalovirus.
OS Synthetic.

XX
FH Location/Qualifiers
FT Complement (3032..4957)
FT CDS
FT /tag- a "EBNA 1 protein"
FT /product- "Epstein Barr Virus Nuclear Antigen 1"
FT /note- "Epstein Barr Virus Nuclear Antigen 1"

XX
XX MO947647-A1.
PD 23-SEP-1999.
XX
PF 12-FEB-1999; 99WO-US03307.
XX
PR 18-MAR-1998; 98US-0040961.
PR 06-AUG-1998; 98US-0130114.
XX
PA (PHAR-) PHARMACOPETIA INC.
XX
PI DamaJ BB, Horlick RA, Robbins AK;
XX
DR WPI; 1999-610610/52.
XX
PT New method for expressing genes from recombinant eukaryotic cells,
PT useful for gene therapy -
XX
PS Example 1; Fig 1; 86pp; English.
XX
XX The present sequence is an anti-sense strand of commercially available
CC plasmid pcwvEBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA
CC 1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain
CC episomes containing EBV origin of replication (oriP) and a gene encoding
CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
CC protein are transfected with these episomes to produce recombinant
CC cell lines expressing multiple genes of interest. This provides a
CC rapid and reliable method of stably expressing multiple genes in
CC transfected cells. The episomes are useful in the transfection of genes
CC encoding receptors, transporter proteins, ion channels, adhesion
CC molecules and transcription factors. The episomes carrying desired genes

CC can also be used to transfect cells in gene therapy, antisense therapy,
CC for gene amplification, cell immortalisation, etc.
XX
SO Sequence 5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;

Query Match 3.6%; Score 57.2; DB 20; Length 5452;
Best Local Similarity 43.2%; Pred. No. 0.023;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

Qy 718 caggcgccactgtgcctccttccttgagagagagctgtgtcgcgagctgcggg 777
Dh 2114 CAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2055
Qy 778 gtgcgagagagtgccctggaacttcatgttcgcgtgtgcgcgcacacacgcgc 837
Dh 2054 GGGCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1995
Qy 838 cagctgtgctcctcgcctcgtgagagcacttcgcgcgcgcgcgcgcgcgcgcgc 897
Dh 1994 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1935
Qy 898 caagacaagctgcgaggttcagagcccaagctctcgcgcgcgcgcgcgcgcgcgc 957
Dh 1934 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1875
Qy 958 aaggtgagagagagagagagagcccgagagcccgagagagagagagagagagtc 1017
Dh 1874 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1815
Qy 1018 accgtgtgag 1077
Dh 1814 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1755
Qy 1078 agcgtgag 1137
Dh 1754 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1695
Qy 1138 cagcgag 1197
Dh 1694 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1635
Qy 1198 ctgagccggcgagggagagcccccacagagagagagagagagagagagagagag 1257
Dh 1634 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1515
Qy 1258 aagatcctcctgaatttgaggggtgtgcccctcagccagagagagagagagagc 1317
Dh 1514 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1515
Qy 1318 caggaagtggcggtcagagaccctggggag 1347
Dh 1514 CAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1485

RESULT 13
Z23778/c
ID Z23778 standard; DNA; 8705 BP.
XX
XX Z23778;
AC
DT 14-JAN-2000 (first entry)
XX
DE Vector pShuttle DNA.
XX
KW Antisense; DNA library; identification; multiple cloning site; MCS;
XX inhibition; ss.
XX
OS Synthetic.
XX
XX WO9950457-A1.
PN
PD 07-OCT-1999.

```

XX 28-MAR-1999; 99WO-US06742.
XX 28-MAR-1998; 98US-0079792.
XX 06-NOV-1998; 98US-0107504.
XX (UTAH ) UNIV UTAH RES FOUND.
XX
XX Rufner DE, Pierce ML, Chen Z;
XX WPI; 1999-610866/52.
XX
XX Production of antisense libraries, used for identifying antisense
XX agents and for identifying target sites for antisense-mediated
XX inhibition of a selected gene
XX
XX Claim 16; Page 43-50; 63pp; English.
XX
XX This invention describes a novel method for generating an antisense
XX library targeted to a selected RNA transcript. The methods can be used
XX for identifying antisense agents and for identifying target sites for
XX antisense-mediated inhibition of a selected gene. The use of a direct
XX library for target site selection significantly simplifies the screening
XX process, since only very small libraries need be prepared and assayed.
XX This sequence represents the vector pshuttle which is used in the method
XX of the invention.
XX
SQ Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other;

```

```

Query Match 3.6%; Score 57.2; DB 20; Length 8705;
Best Local Similarity 43.2%; Pred. No. 0.024;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

```

```

OY 718 cagcgccacactgtcgccttctcttgagagacgtgtgctgagcgagctgcggg 777
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7979 CAGGAGCAGAGAGAGGCGCAGAGCAGAGAGGCGCAGAGAGGCGCAGAGAG 7920
OY 778 gtcggcgagagtgccctgactcactcactgttcgcgcgtgctgcgcacacagcgcc 837
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7919 GGGCAGAGCAGAGAGAGGCGCAGAGCAGAGAGGCGCAGAGAGGCGCAGAG 7860
OY 838 cagctgtgtcacttcgcctgtgagacacttcgcgcgcgcctgcgaattcgtctggggccc 897
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7859 GAGCAGAGAGAGGCGCAGAGCAGAGAGGCGCAGAGAGGCGCAGAGAGAGG 7800
OY 888 caagacaagctgcggaggttcaagccagctctctgcgcgacatcgcctgagaggtctcagg 957
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7799 CAGGAGGCGCAGAGAGGCGCAGAGCAGAGAGGCGCAGAGAGGCGCAGAG 7740
OY 958 aaggtcgagagagagagagagagagagagagagagagagagagagagagagag 1017
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7739 GGGCAGAGCAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGAG 7680
OY 1018 acctgtgagcagacatagcaagagtggtgggcaaggtgcagagccagccagcg 1077
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7679 GAGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGAG 7620
OY 1078 agctgtgagagagagagagagagagagagagagagagagagagagagagag 1137
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7619 GGGCAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGAG 7560
OY 1138 cagcgagagagagagagagagagagagagagagagagagagagagagagag 1197
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7559 GAGGCGCAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGAGAG 7500
OY 1198 ctgagcgagcgagagcagcgcccaagagccctcagagtgctcctcagaggtgag 1257
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7499 CAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGAGAG 7440
OY 1258 aagatcgtcgaatttgagaggtgtgtgcctcagcagcgagcagcctcagagcgagac 1317
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7439 GGGCAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGG 7380

```

```

OY 1318 caggaagtggcggtcagagacctggggag 1347
   ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 7379 CAGGAGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAG 7350

RESULT 14
ID V21683 standard; DNA; 9600 BP.
XX V21683;
XX
XX 17-AUG-1998 (first entry)
XX
DE Vector plasmid pcMVknTR-EPI.
XX
XX Polynucleotide delivery; plasmid pcMVknTR-EPI; vector;
XX gene therapy; vaccine; polycationic agent; ss.
XX
XX Chimeric - Epstein-Barr virus.
XX OS Chimeric - Adeno-associated virus.
XX OS Chimeric - Cytomegalovirus.
XX OS Chimeric - Bos taurus.
XX
XX Key Location/Qualifiers
XX CDS 14..2594
XX
XX /*tag- a
XX /*product- "EBV nuclear antigen A"
XX 2623..4559
XX /*tag- b
XX /*note- "EBV origin of replication"
XX 4928..5104
XX /*tag- c
XX /*rpt-type- INVERTED
XX /*note- "AAV inverted terminal repeat"
XX 7189..7355
XX /*tag- d
XX /*rpt-type- INVERTED
XX /*note- "AAV inverted terminal repeat"
XX 5112..6734
XX /*tag- e
XX /*note- "CMV immediate-early enhancer/promoter"
XX 6818..7050
XX /*tag- f
XX /*note- "bovine growth hormone polyA sequence"

XX WO9806437-A2.
XX
XX 19-FEB-1998.
XX
XX 13-AUG-1997; 97WO-US14465.
XX
XX 13-AUG-1996; 96US-0023867.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Cohen F, Dubois-Stringfellow N, Dwariki V, Innis MA;
XX Murphy JE, Tetsuo U, Zukermann R;
XX WPI; 1998-159296/14.
XX
XX Polycationic agents based on alpha-amino acids, able to complex
XX with nucleic acid - to facilitate its entry into cell, condense it
XX and protect it against serum degradation, particularly for use in
XX gene therapy
XX
XX Disclosure; Page 77-80; 100pp; English.
XX
XX This polynucleotide comprises the DNA sequence of vector plasmid
XX pcMVknTR-EPI, which contains an Epstein-Barr virus (EBV) origin
XX of replication from plasmid pCEP4, a coding region for EBV nuclear
XX antigen A from pCEP4, a pair of inverted terminal repeats from
XX adeno-associated virus, a cytomegalovirus enhancer/promoter, a

```

CC bovine growth hormone polyA sequence, and a kanamycin resistance
CC selectable marker. Polynucleotides encoding polypeptides, such as
CC erythropoietin or leptin, and ribozymes and antisense
CC polynucleotides can be inserted into the vector. The vector is
CC preferred for use in novel compositions and methods for improved
CC polynucleotide delivery into cells. In these methods, polycationic
CC agents are used to increase the frequency of uptake of a
CC nucleic acid (see also V21684-86) into a cell. The polycationic
CC agent can condense with the nucleic acid and inhibit serum and/or
CC nuclease degradation of the nucleic acid. The nucleic acid can be
CC a vector, may express a therapeutic protein or a vaccinating viral
CC or cancer antigen, or is itself therapeutic (antisense or
CC ribozyme). The methods and compositions can be used in the gene
CC therapy of many diseases.

SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 3.6%; Score 57.2; DB 19; Length 9600;

Best Local Similarity 43.2%; Pred. No. 0.024;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

```
QY 718 caggcgccactgtccgtcttccttccttgagagacgctgtgctgctgagctgcgggg 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 737 caggagcagagagagagagagagagagagagagagagagagagagagagagag 796
QY 778 gtgcgagagagagagagagagagagagagagagagagagagagagagagag 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 797 gggcagagagagagagagagagagagagagagagagagagagagagagag 856
QY 838 cagctgtgctccttcgctgagacatctccgcccgcctcaagctgtctggtggccc 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 857 gaggcagagagagagagagagagagagagagagagagagagagagagagag 916
QY 898 caagacaagctgcgagaggttcacagccctctgcctgcctcgcctgcagagctccag 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 917 caggagagagagagagagagagagagagagagagagagagagagagagag 976
QY 958 aagctggagagagagagagagagagagagagagagagagagagagagagag 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 977 gggcagagagagagagagagagagagagagagagagagagagagagagag 1036
QY 1018 acctgtgagcagagacatagaaaggttgagagagagagagagagagagagag 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1037 gaggcagagagagagagagagagagagagagagagagagagagagagagag 1096
QY 1078 agcgttgagagagagagagagagagagagagagagagagagagagagagag 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1097 gggcagagagagagagagagagagagagagagagagagagagagagagag 1156
QY 1138 cagcggagagagagagagagagagagagagagagagagagagagagagag 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1157 gaggagagagagagagagagagagagagagagagagagagagagagagag 1216
QY 1198 ctgagcggcagcagagagagagagagagagagagagagagagagagagag 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1217 caggagagagagagagagagagagagagagagagagagagagagagagag 1276
QY 1258 aagctgcttgaatttgagagaggtgtgctctcagcagagagagagagagag 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1277 gggcagagagagagagagagagagagagagagagagagagagagagagag 1336
QY 1318 caggagagagagagagagagagagagagagagagagagagagagagagag 1347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1337 caggagagagagagagagagagagagagagagagagagagagagagagag 1366
```

RESULT 15

Z22248

ID Z22248 standard; DNA; 10380 BP.

XX Z22248;

AC

XX

DT 29-NOV-1999 (first entry)

XX Nucleotide sequence of pCEP4 vector.

DE G-1 protein coupled receptor; G-1PCR; G-1a protein; cytoplasmic calcium;

KW calcium mobilization; nociceptin receptor; chemokine receptor; CCR3;

XX CCR2; interleukin 8 receptor type B; episomal expression vector; ss.

OS Synthetic.

XX WO99/47921-A1.

XX 23-SEP-1999.

XX 09-FEB-1999; 99WO-US02852.

XX 18-MAR-1998; 98US-0040961.

XX (PHAR-) PHARMACOPERA INC.

XX Dama J BB, Horlick RA, Robbins AK;

XX A method to enhance signal transduction responses involving cytoplasmic

XX calcium release

XX Examples; Fig 3a-J; 64pp; English.

XX The invention relates to a cell transfected with a gene encoding a G-1

XX protein coupled receptor (G-1PCR) and with a gene coding for a G-1a

XX protein capable of coupling to the G-1PCR, to increase the cytoplasmic

XX calcium upon binding of an agonist to the G-1PCR. The cell is used to

XX assay a test compound for its effect on a G-1PCR. The method and cells

XX are useful for assaying agonist and antagonist compounds, which bind to

XX G-1PCR which, stimulate intracellular calcium mobilization, such as the

XX nociceptin receptor, and particularly chemokine receptors such as CCR3,

XX CCR2 and the interleukin 8 receptor type B. The present sequence

XX represents the nucleotide sequence of the pCEP4 vector.

XX Sequence 10380 BP; 2560 A; 2561 C; 2964 G; 2295 T; 0 other;

Query Match 3.6%; Score 57.2; DB 20; Length 10380;

Best Local Similarity 43.2%; Pred. No. 0.024;
Matches 272; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

```
QY 718 caggcgccactgtccgtcttccttccttgagagacgctgtgctgctgagctgcgggg 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6223 caggagcagagagagagagagagagagagagagagagagagagagagagagag 6282
QY 778 gtgcgagagagagagagagagagagagagagagagagagagagagagagag 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6283 gggcagagagagagagagagagagagagagagagagagagagagagagag 6342
QY 838 cagctgtgctccttcgctgagacatctccgcccgcctcaagctgtcgtctggggccc 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6343 gaggcagagagagagagagagagagagagagagagagagagagagagagag 6402
QY 898 caagacaagctgcgagaggttcacagccctctgcctgcctcgcctgcagagctccag 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6403 caggagagagagagagagagagagagagagagagagagagagagagagag 6462
QY 958 aagctgagagagagagagagagagagagagagagagagagagagagagagag 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6463 gggcagagagagagagagagagagagagagagagagagagagagagagag 6522
QY 1018 acctgtgagcagagacatagaaaggttgagagagagagagagagagagagagag 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6523 gaggcagagagagagagagagagagagagagagagagagagagagagagag 6582
QY 1078 agcgtgagagagagagagagagagagagagagagagagagagagagagagag 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Mon Oct 16 10:08:49 2000

us-09-431-843b-7.rng

Page 12

[illegible]

Search completed: October 12, 2000, 18:05:10
Job time: 30768 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using 'sw model

Run on: October 12, 2000, 12:02:37 ; Search time 7878.83 Seconds
(without alignments)
1256.363 Million cell updates/sec

Title: US-09-431-843B-7

Perfect score: 1601

Sequence: 1 tagaattcgcgcgcgtga.....ccctctcgtgagtgggggcg 1601

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*\n2: gb_est2:*\n3: gb_est3:*\n4: gb_est4:*\n5: gb_est5:*\n6: gb_est6:*\n7: gb_est7:*\n8: gb_est8:*\n9: gb_est9:*\n10: gb_est10:*\n11: gb_est11:*\n12: gb_est12:*\n13: gb_est13:*\n14: gb_est14:*\n15: gb_est15:*\n16: gb_est16:*\n17: gb_est17:*\n18: gb_est18:*\n19: gb_est19:*\n20: gb_est20:*\n21: gb_est21:*\n22: gb_est22:*\n23: gb_est23:*\n24: gb_est24:*\n25: gb_est25:*\n26: gb_est26:*\n27: gb_est27:*\n28: gb_est28:*\n29: gb_est29:*\n30: gb_est30:*\n31: gb_est31:*\n32: gb_est32:*\n33: gb_est33:*\n34: gb_est34:*\n35: gb_est35:*\n36: gb_est36:*\n37: gb_est37:*\n38: gb_est38:*\n39: gb_est39:*\n40: gb_est40:*\n41: em_estda:*\n42: em_estfun:*\n43: em_esthum1:*

44: em_esthum2:*\n45: em_esthum3:*\n46: em_esthum4:*\n47: em_esthum5:*\n48: em_esthum6:*\n49: em_esthum7:*\n50: em_esthum8:*\n51: em_esthum9:*\n52: em_esthum10:*\n53: em_esthum11:*\n54: em_esthum12:*\n55: em_esthum13:*\n56: em_esthum14:*\n57: em_esthum15:*\n58: em_esthum16:*\n59: em_esthum17:*\n60: em_esthum18:*\n61: em_esthum19:*\n62: em_esthum20:*\n63: em_estin1:*\n64: em_estin2:*\n65: em_estin3:*\n66: em_estin4:*\n67: em_estom:*\n68: em_estov1:*\n69: em_estov2:*\n70: em_estpl1:*\n71: em_estpl2:*\n72: em_estpl3:*\n73: em_estpl4:*\n74: em_estpl5:*\n75: em_estpl6:*\n76: em_estro2:*\n77: em_estro3:*\n78: em_estro4:*\n79: em_estro5:*\n80: em_estro6:*\n81: em_estro7:*\n82: em_estro8:*\n83: em_estro9:*\n84: em_estro10:*\n85: em_estro11:*\n86: em_estro12:*\n87: em_estro13:*\n88: gb_gss1:*\n89: gb_gss2:*\n90: gb_gss3:*\n91: gb_gss4:*\n92: em_gss1:*\n93: em_gss2:*\n94: em_gss3:*\n95: em_gss4:*\n96: gb_gss15:*\n97: gb_gss16:*\n98: gb_gss17:*\n99: gb_gss18:*\n100: gb_gss19:*\n101: em_gss5:*\n102: em_gss6:*\n103: em_gss7:*\n104: em_gss8:*\n105: em_gss9:*\n106: em_gss10:*\n107: em_gss11:*\n108: em_gss12:*\n109: em_gss13:*\n110: gb_gss5:*\n111: gb_gss6:*\n112: gb_gss7:*\n113: gb_gss8:*\n114: gb_gss9:*\n115: gb_gss10:*\n116: gb_gss11:*

	Db	421	TGAGGGCTCCAGGAAGGTGGAGGAGAAGAAACCCC GGSGAACCCCGAACCAGAGGCCA	480
Oy	1004	gcaaccaaggtctcgacctgtggccaaagaatagaaggatgtaggcagagtggggcaggttgacg-ag	1062	
Dd	481	GCAACCAGAGGTCGCACCTC GTGGCCAGAGCATATACAAGGGTGGGGCGAGGGTGACCAAG	540	
Oy	1063	ggsgcccacgaccagagaggttggagcccccaagatatgcgagacccttcggaggaagccagggg	1122	
Dd	541	GGGGCCCCAGCACAGGAGGCTGGAGCCCCCAGATATCCG GAGACTCCCTCGAGAGAGCCA-GGG	599	
Oy	1123	gatggcagagggggccccaagggaatagagcccgagaccctaagccccaagaagagacaag	1182	
Db	600	GATGAGGCAAGGGGGCCACGGGGAAGATAGCCCGGAGCCCTTAAGGCCCAAAGAGAGCAAG	659	
Oy	1183	aagaagaaactcygactgtaagccggcgaggaagcacgcgccacaag	1225	
Db	660	- AGAGGAAGCTGAGAGCTGCGGGGGAGACACGCCCAAAAGAG	701	
RESULT	2			
BE277227				
LOCUS	BE277227	868 bp mRNA EST	13-JUL-2000	
DEFINITION	60117859.FP1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3050950 5'			
VERSION	mRNA sequence.			
KEYWORDS	BE277227 BE277227.1 GI:9152194			
SOURCE	EST.			
ORGANISM	human.			
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	N1 (bases 1 to 868)			
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: ATCC/DCTD/ CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LHCN97 row: h column: 23 High quality sequence stop: 742.			
FEATURES	Location/Qualifiers			
Source	1..868 /organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:3050950"			
	/clone_lib="NH MGC.20"			
	//ssue_type="melanotic melanoma"			
	//lab_host=DHI0B (phage-resistant)"			
	/note="Organ: skin; Vector: pOTR7; Site_1: XhoI; Site_2:"			
	EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald W. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." "			
BASE COUNT	171 a 245 c 294 g 157 t 1 others			
ORIGIN				
Query Match	37.3%; Score 596.6; DB 34; Length 868:			
Best Local Similarity	88.1%; Pred. No. 5.4e-113;			
Matches 760; Conservative 0; Mismatches 74; Indels 29; Gaps				
44 ccgaactgcgaactcgaactcgggagaggaagcatcggaagacgggaggaagag				

Db	1	CGAGCTGGACTCCACCTGGGAGGAGGACGAGGAGTACGGAGACGCGGAGGACGAG	60
Qy	104	actgc-gagagccgagagcc-----gccgcgcgaggaagcgcgagc	145
Db	61	ACTGCTGGAGACGGCGAGTGGCTCTCTGGCGCTAGGGTGAATGTGTACTGCAAGGGTGAC	120
Qy	146	caggggagaggaagagagagtcgtaggaagccgcggggcgcgccgaagtcgcttcagt	205
Db	121	TGAGGACGATGAGTGTGGATGTAGTATGCTGTGTGGCGCTGCGGCCACAGCTCGTTCCAGT	180
Qy	206	ccagatctacagaggttccgaagac-tggcgagccacgagggagacatgttgaatctcggcac	264
Db	181	CCAGATGTACAGGGGTCCAGAAACTTGGCGAGCCACGAGGACATGTATAGTATCGGCAC	240
Qy	265	aactaccgcgtctgtgtgaacagagactgcaatgtggygacacgcgcaaaccttgagtttac	324
Db	241	AACATATCCGGATCTGTGTGAGACGAGACTGCATATGGGGACAGCCAAACTGTAGTTTAC	300
Qy	325	agaaatcgatccgccttcctcgcgaagagctgtttcaattgagacatcttcagaacttg	384
Db	301	AGAAATGAGATCCCTCTCTCCCAAGGCGCTTTCATGTGGGCATCTTCAGAACTGG	360
Qy	385	acggacaactctgaccttcctcttgaggaaatcacatccctcacatccagttgcgtttccctg	444
Db	361	ACGGACACTTGTATGACTCTTTGAGAGCAATACCTCTACATCCAGTGGCTTTTCTCTG	420
Qy	445	cgaagaccagagagtgaaactgacatgcgcaagacccttcacgctcagagagagtcgagtg	504
Db	421	CGAGAACCGAGAGTGAATGCTGCAATGCGCAAGCCCTCAAGCTCAGAGGAGTGTGAGGTTT	480
Qy	505	aaaagctccagagatccagagagcgtgtgtccgggacctagagctcatgtctggtcttc	564
Db	481	AAAAGCTCCAGGAGATCCAGAGACGGCTTGTCCGGGCTACGAGCTCATGCTGGCTTC	540
Qy	565	tac-ggagtcgggtcttgagagacgagagcagcgagcaggttggccgagacaagaactaca	623
Db	541	TACGGGATCCGGCTGGAGGACCGAGGACGGGACGGGTGGGCGGACACAGACTACCA	600
Qy	624	gaagcgcctccagaacacttgaactgtgcgcagcacaacaacactccgcacatcacgcatcct	683
Db	601	GAGGCGTTCAGAACTGAACTGGCGAGCCAGCAACAACCTCCGATCACAGCATCTCT	660
Qy	684	caagtcgcgctgtgagctgagctcgcagacacttccag--cgccactgtgtcgtcttc	741
Db	661	CAAGCTGCTGGGTAGAGCTGGGCTCTCAGCACTTCCAGGGGCGCCGCTGGCTCTTTC	720
Qy	742	ctgga-ggagacgctgtgtgtg--cggagagtcggcgggg--tgcggcagagtgcccttg	796
Db	721	CTGGAGGGAGACGCTGTGTGCGGCGGTGACGCTGCGGGGGGTGCGGACAGAGTGCCTTG	780
Qy	797	--actactcatgtctgcgcgtgcgtgcgcagacacagcgcgcgcagctgtgtcacttcgc	854
Db	781	GACTTACTTCAATGGTCCGCCGTGCGCTGCCACACACAGGCTGCAGCTGTTGCCCTCGC	840
Qy	855	ctggagagcacttcggccgcgt 877	
Db	841	CTTGGAGACTTTCGGGCGCCGCT 863	
RESULT	3		
BE387193			
LOCUS	BE387193	751 bp	mrna
DEFINITION	60127706F1 NIH_MGC_20	Homo sapiens	cdna clone IMAGE:3618136 5',
ACCESSION	BE387193		
VERSION	BE387193.1	GI:9332558	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
AUTHORS	Mammalia: Euthera: Primates: Catarrhini: Hominiidae: Homo.		
	1 (bases 1 to 751)		
	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .		

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCID/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LCM286 row: 1 column: 17
High quality sequence start: 86
High quality sequence stop: 727.
Location/Qualifiers
1.751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3618136"
/clone.lib="NIH.MGC.20"
/tissue.type="melanotic melanoma"
/lab.host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 170 a 202 c 254 g 125 t
ORIGIN
Query Match 37.0%; Score 592.6; DB 35; Length 751;
Best Local Similarity 96.8%; Pred. No. 3.5e-11;
Matches 626; Conservative 0; Mismatches 19; Indels 2; Gaps 2;
Db 10 GCGGCGCTGTAATCTAGCGAGCATGAGACCGCGGAGTGTGACCTGACCTGAGAGAG 69
38 GCGAGGCGCGGATGTCTGCTACTATGTGACCTCGACTCGGACTTCCACTCGGAGAGAG 97
Qy 70 GAGCGAGGAGATGCGAGGAGCGCGAGGAGCGAGGAGCGGAGCGGAGCGGAGCGG 128
Db 98 GAGCGAGGAGATGCGAGGAGCGCGAGGAGCGAGGAGCGGAGCGGAGCGGAGCGG 157
Qy 129 GCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAG 188
Db 158 GCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAG 217
Qy 189 GCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAG 248
Db 218 GCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAGGAGCGGAG 277
Qy 249 GTGTAGGATCGGACACTATCCGATCTGTTGAGACGAGACTCAATGGGAGACCGC 308
Db 278 GTGTAGGATCGGACACTATCCGATCTGTTGAGACGAGACTCAATGGGAGACCGC 337
Qy 309 AAACCTGAGTTCTCAAGAAATGAGATCCGCTCTTCCCAAGGCTGTTCATTGAGGA 368
Db 338 AAACCTGAGTTCTCAAGAAATGAGATCCGCTCTTCCCAAGGCTGTTCATTGAGGA 397
Qy 369 CATCTTCAAGAACTGAGACAACTATGACCTCTTAGAGGACAACTCACTCTCAATCA 428
Db 398 CATCTTCAAGAACTGAGACAACTATGACCTCTTAGAGGACAACTCACTCTCAATCA 457
Qy 429 GTGTAGGATCGGACACTATCCGATCTGTTGAGACGAGACTCAATGGGAGACCGC 488
Db 458 GTGTAGGATCGGACACTATCCGATCTGTTGAGACGAGACTCAATGGGAGACCGC 517
Qy 489 GAGAGTCAGGTCGTTAAAGCTCCACAGAGATCCAGAGCGGCTTGTCCGCGCTACGA 548
Db 518 GAGAGTCAGGTCGTTAAAGCTCCACAGAGATCCAGAGCGGCTTGTCCGCGCTACGA 577

Qy 549 gctcatgctggctctctacgagatccgctgagagaccagagcagggacggt-99gcc 607
Db 578 GCTCATGCTGGGCTTCTACGGGATCCGCTGAGAGACGAGGACGCGGCGTGGGGCC 637
Qy 608 gagcagaaactacagagagcgcttcagaaactcgaactcggcagc 654
Db 638 GAGCAGAACTACAGAGAGCGCTTCAGAACTGAACTGGCGAGC 684
RESULT 4
BE385629 682 bp mRNA EST 21-JUL-2000
LOCUS BE385629
DEFINITION 601278145F1 NIH.MGC.20 Homo sapiens cDNA clone IMAGE:3618922 5',
mRNA sequence.
ACCESSION BE385629
VERSION BE385629.1 GI:9330994
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 682)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCID/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LCM286 row: 1 column: 11
High quality sequence start: 20
High quality sequence stop: 663.
Location/Qualifiers
1.682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3618922"
/clone.lib="NIH.MGC.20"
/tissue.type="melanotic melanoma"
/lab.host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 160 a 190 c 224 g 108 t
ORIGIN
Query Match 36.6%; Score 586; DB 35; Length 682;
Best Local Similarity 96.2%; Pred. No. 8e-11;
Matches 654; Conservative 0; Mismatches 20; Indels 6; Gaps 5;
Db 40 GACCCGAGTGCAGATCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 98
3 GACCCGAGTGCAGATCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 62
Qy 99 GAGAGATCGGAG 158
Db 63 CGAGAGATCGGAG 122
Qy 159 CGAGAGATCGGAG 218
Db 123 CGAGAGATCGGAG 182

|||||
Db 452 AAGATAGCCGAGCCCTTAAGCCCAAGAGACAGAGAGACTGGAGTGAGCC 511
LOCUS
QY 1205 ggcggagcagcgcgc 1220
|||||
Db 512 GCGCGGAGCAGCCGCC 527
*
RESULT 8
BE409569 1296 bp mRNA EST 21-JUL-2000
LOCUS 601299883F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629948 5',
DEFINITION mRNA sequence.
ACCESSION BE409569
VERSION BE409569.1 GI:9346019
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1296)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM217 row: e column: 21
High quality sequence start: 2
High quality sequence stop: 532.
Location/Qualifiers
1. 1296
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629948"
/clone_1kb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 327 a 416 c 389 g 164 t
ORIGIN
Query Match 27.68; Score 442; DB 35; Length 1296;
Best Local Similarity 84.68; Pred. No. 3,4e-81;
Matches 668; Conservative 0; Mismatches 90; Indels 32; Gaps 14;
QY 515 agggatccaggaac-ggctgtccggcgctcagagtcagtcgggctctcaggaac 573
|||||
Db 11 AGGGATCCAGGAGCTAGCTGTGTCGGGCTTACAGCTCATGCTGGGCTTCTACGGGATC 70
QY 574 cgcgttgagagcagcagcagcagc-gcacggtgagcgagcagcagcagcagcgtt 632
|||||
Db 71 CGGCTGGAGGAGCAGCAGCAGCGGCGCAGGTGGCGGAGCAGAACTACAGAGCGCTT 130
QY 633 ccgaacctgaactgtgcgcagcacaacaacctccgcatcagcagcagcagcagcgtcc 692
|||||
Db 131 CCAAGAACTGAACTGGCGCAGCAGCAACCAACCTCCGATCATCAGCATCTCCCAAGTGCT 190
QY 693 gtgagactgagctcgagacattccagcgagcagcgttcgctcttctcctggagagac 752

|||||
Db 191 GGGTAGACTGGGCTCGAGAGACTTCCAGGCGCGCGTGTGCTCGCTTCTCTGAGAGAC 250
QY 753 gctgtgtcggcggagagctgcgggggtgtgcgcaagatgcccctgaactcattc-tgc-g 811
|||||
Db 251 GCTGTGTCGGCGGGAGAGCTGCGGGGGGTGCGCGAGAGTGCCCTGGACTTCACTGTTCA 310
QY 812 ccgttgagctgcgcaaccagc-gcgcagagctgtgctactgcctcgtggagactccgg 870
|||||
Db 311 CCGTGCCTCCGACACACACACACACAGCTGTGACTTCCCTGGAGCACTTCGCG 370
QY 871 cccgcg-tgcaatgctg--ctgaggcccccaagacaagctgtggagaggttcaagccagc 927
|||||
Db 371 ACCCGCATGCAAGTCTCTATGGGGGCCCAAGACAGCAAGCTGCGGAGCTCAAGCCAGC 430
QY 928 tctctgcgcat-cgcgtcagggctccagaaagtgtgagagagaaagcccgagg 966
|||||
Db 431 TCTCTGCCCATACCGCTCGAGGCTTCCAGGAAGTGGAGAGGAAGGAGCCCGGGA 490
QY 987 cccgcagcagagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1040
|||||
Db 491 CCGGACACAGAGGCGCAGCAGCGCTGCGACCTGTGGCCAGAGCCATAGCAAGGCT 550
QY 1041 gggctggggaggggtgagcagagggccccagcagcagcagcagcagcagcagcagc 1097
|||||
Db 551 GCGGGCAGCAGTGGACAGACGGGACCCAGCCAGCGAGCTGAGCCCAAGAGATGCC 610
QY 1098 gggacccttgagagagagcagcaggggga--tgaggcagggggccagcgggagagatagcc 1154
|||||
Db 611 GGAGCACTACAGAGGAGCAGCGGAGCTGAGCGCAGCGGCGCAGCGGCAAGCATCA 670
QY 1155 gggg-----cccttaagcccaagagagcagcagagagagagagcttga----gctagcc 1204
|||||
Db 671 GCAAGAGCCCATATACCCAGCAAGAGAGCAGCAGCAGCAAGAGTGGCGCTGACC 730
QY 1205 ggcggagcagcgcgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 1264
|||||
Db 731 CCGGAGAGCATCCG-CCACAGAGCCCGAGCTCAGAGTGCCTCA-AGGTGCAAGACAGC 788
QY 1265 ctctgaattt 1274
|||||
Db 789 CACTGAAGCTT 798
RESULT 9
BE391750 757 bp mRNA EST 21-JUL-2000
LOCUS 601283311F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3604994 5',
DEFINITION mRNA sequence.
ACCESSION BE391750
VERSION BE391750.1 GI:9337115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 757)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM252 row: f column: 03
High quality sequence start: 110
High quality sequence stop: 686.

OY 327 aaatagatcgccttcctcgcacagcgtgttcattgagagacattcttgaactgac 386
|||||
DB 305 AATATAGATCCCTTCCTCCGACAGCGCTTTCATGAGCATTCCTCAGACTGGAC 364
|||||
OY 387 ggaacacatgacccctcttgagacatcctcaccagtggtgttccctcgtg 446
|||||
DB 365 GGACACATATGACCTCCTTGAGACATCATCCTACATCCAGTGGCTTCTCTGCG 424
|||||
OY 447 agaacacag 455
|||||
DB 425 AGAACGAG 433
|||||

RESULT 11
A1816189/c 432 bp mRNA EST 09-JUL-1999
LOCUS au45h07.x1 schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION IMAGE:251757 3', mRNA sequence.
A1816189
ACCESSION A1816189.1 GI:5431735
VERSION EST.
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,
Kirizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,
J., Moore,B., Schellenberg,K., Stepcoe,M., Tan,F., Theisling,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Other-ESTs: au45h07.y1
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 431.
Location/Qualifiers

FEATURES

Source

1..432

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:251757"

/clone_lib="Schneider fetal brain 00004"

/sex="male"

/tissue_type="frontal lobe"

/dev_stage="5 months post-conception"

/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGATCCTTATTAATTAATATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTCAAGTCTTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."

BASE COUNT 73 a 143 c 142 g 73 t 1 others
ORIGIN

Query Match

26.4%; Score 423; DB 13; Length 432;

Best Local Similarity 98.6%; Pred. No. 2,6e-77;
Matches 426; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 467 atgccaagccctcaccagctcaggagagcgtgtttaaagctcccaaggatccagg 526
|||||
DB 432 ATGCCAACCCCTCAGCTCAGGAGGTCGAGGTGTTAAAGCTCNCAGAGATCCAGG 373
|||||
OY 527 agcgcttgctcgggctcaggatcagctatgctgggtcttaaggatccgcttgaggacc 586
|||||
DB 372 AGCGGCTTGCTCGGGCTCAGAGCTCATGCGGGCTTCTAGGGATCCGCTGAGAGACC 313
|||||
OY 587 gaggcagcgacagctgtggcgccagacagaaactaccagacgcttcagacactgaact 646
|||||
DB 312 GAGCAGCGGACGAGCTGGGCGCTGTGCACAGAACTACAGAAACGCTTCCAGAACTGAAGT 253
|||||
OY 647 ggcgagccacaacaacccctcgcacacacgcatcctcaactcgcgtgtgagctgagcc 706
|||||
DB 252 GCGCAGCCACACAACTCCGACATCACACGATCCTCAAGTGGCTGAGTGGGCGC 193
|||||
OY 707 tcgagcacttcagagcgccactgtgtcgcctctctctctggaagagagcgtgtgagcgagg 766
|||||
DB 192 TCGAGCACTTCAGCGCGCGCGCTGCTCTCTCTGAGAGAGACGCTGCTGGCGCGG 133
|||||
OY 767 agctgcgggggtgtgcgagagtgccctggaactcatgttcgcgtgcgtgcgcgac 826
|||||
DB 132 AGCTCCCGGGGCTGGCGAGAGTGGCTCTGACTTATCTTCCCTGCGCTGCGGAC 73
|||||
OY 827 accaagcgccgcagctgtgtcacttcgccttgaggagacattccgcccgcgtcgaattcg 886
|||||
DB 72 ACCACGCGCGCAGCTGTGCTGACTTCGCTGGAGCACTTCGCGCGCGCTGCAAGTTGCG 13
|||||
OY 887 tctggggggcccc 898
|||||
DB 12 TCTGGGGGCCCC 1
|||||

RESULT 12

LOCUS

AM250708

DEFINITION

AM250708

ACCESSION

AM250708

VERSION

AM250708.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other-ESTs: 2822396.3prime
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTF cDNA Library Preparation: Ling
Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LNL) DNA Sequencing by: Berkeley MGC sequencing
project clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/dbp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross-match from University of Washington Genome Center
PHRAP suite. Poly-T identification: patMatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:
http://www.genome.washington.edu
Plate: L10M9 row: E column: 21
High quality sequence stop: 183.
Location/Qualifiers

FEATURES

Source

1..432

/organism="Homo sapiens"

/db_xref="taxon:9606"

```

/clone="IMAGE:2822396"
/clone_lib="NIH-MGC-7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Lung; Vector: pORF1; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      80 a      141 c      143 g      68 t
ORIGIN
```

```

Query Match      25.9%; Score 415; DB 21; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.2e-75;
Matches 418; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```

QY 505 aaagctcccaagagatcccaagagagctgtccgggctcaagatcgtggcttc 564
    |||||||
Db 10 AAAAGCTCCAGAGATCCAGAGCGGCTTGTCCGGCTACAGCTCATGCTGGGCTTC 69
    |||||||
QY 565 tacggatccggctgtgagacccagagcgagctgtggcgagacagaaatccag 624
    |||||||
Db 70 TAGGGGATCCGGCTGGAGGACCGAGCACGGGACGGTGGCCGAGACAGAACTACCA 129
    |||||||
QY 625 aagcgctccagaacctgaactgtgcgacgacacaaacacctccgcatcacagcatc 684
    |||||||
Db 130 AACCGCTCCAGAACCTGAACCTGGCGACGACACAACTCCGATACAGCATTC 189
    |||||||
QY 685 aagtcgcgtgtgagctgagacccctcagacccctcagacccctcgtcttcctg 744
    |||||||
Db 190 AATTCCTGGGTAGCTAGGCTTGAAGCTTCAAGCGCCACTGTGCTGCTTCTTC 249
    |||||||
QY 745 gaggaaagcgtgtgtgcgagcgagctgcgaggtgtgcgacagatcccttgactatc 804
    |||||||
Db 250 GAGGAGACGCTGTGCGGGGAGTGTCCGGGGGTGCGGAGAGTCCCTGGACATCTTC 309
    |||||||
QY 805 atgttcgctgtgtgtgcgacacacagcgccgacagctgtgcaattcgctgtggagac 864
    |||||||
Db 310 ATGTTCGCTGTGCTGTGCGGACACAGCGCCGCACTGTGCTGCTGCGGAGCAC 369
    |||||||
QY 865 ttccggcccgctgtcaagttcgtctgtggggccccaagaaagctgtggagttcaagccc 924
    |||||||
Db 370 TTCGGCCCCCGCTGCAAGTTCGTCTGTGGGGCCCAAGACAGAGCTGGAGGTTCAAGCCC 429
    |||||||
QY 925 agc 927
    |||
Db 430 AGC 432
```

```

RESULT 13
LOCUS BE282484 654 bp mRNA EST 13-JUL-2000
DEFINITION 601103075F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3495186 5',
mRNA sequence.
ACCESSION BE282484
VERSION BE282484.1 GI:9158085
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```

REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
AUTHORS 1 (bases 1 to 654)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
```

```

FEATURES
source
1..654
/organism="Mus musculus"
/strain="C57BL/6J (F1)"
/db_xref="taxon:10090"
/clone_lib="NIH-MGC-7"
/clone="IMAGE:3495186"
/clone_lib="NCI CGAP Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/notes="Organ: Lung; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
```

```

Query Match      25.0%; Score 400.8; DB 34; Length 654;
Best Local Similarity 78.4%; Pred. No. 9.9e-73;
Matches 480; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
```

```

BASE COUNT      179 a      175 c      181 g      119 t
ORIGIN
```

```

QY 93 ggaagacagagatcgtcgaagagcgagcgccgagcgagcgagcgagcgagcgag 152
    |||||||
Db 14 GGAGACGAGAGCGAGAGAGCGGCGAGCATGGCCAGGCGAGATGACAGCATGAGGA 73
    |||||||
QY 153 cgaggaagagatcgtcgaagagcgagcgcgagcgcgagcgagcgagcgagcgag 212
    |||||||
Db 74 CACGGGCGACATACCGGCGAGGAGGACCGGCGAGGCGGCGAGGCGGCGAGGCGG 133
    |||||||
QY 213 gacagagtcgaagacgtgcgagcgaagagagcatgtgagttatcgagacatctc 272
    |||||||
Db 134 GACAAAGTACGGAATCTGGCGTGTATGACAGGACATGCAAAAGATACCGGACACTACC 193
    |||||||
QY 273 ggaatcgtgtggaagagacgtgcaatgtggagacgccaacgtgagttctacagaaatga 332
    |||||||
Db 194 GGATTTGACAGATCAAGACATGCAAGGAGCATGTGCACACTGAGCTGTACAAAATGA 253
    |||||||
QY 333 gatcgcctcctgcgaagcgctgtttcatgtgagacatcttcagaactgtgcgagaa 392
    |||||||
Db 254 GATCTGCTTCAGGCAAAATGCTTTCATTTAGGACATTTTCAAGACTGGAAAGACA 313
    |||||||
QY 393 ctatgacctccttgagagacaatactactacatcagtagtctgttctcctgcgagaa 452
    |||||||
Db 314 CTATGACCTCTGGAAGAGATCATCTCATCATCAGTGGCTGTTCCCTGAGGGAAC 373
    |||||||
QY 453 aggaatgaactgcatgccaagccctcaagcgtcagagaggtgcagagttttaaagctc 512
    |||||||
Db 374 AGGAGTGAATGCAATGCAAGGCGCTTCACTGAGAGGAGGTTGAGGCACTTTAAAGCTC 433
    |||||||
QY 513 ccaggaatccagagcgagctgtccgggctcagagacatgctgtgtgagagagag 572
    |||||||
Db 434 CAAGGAAGTCAGAGAGCGTCTTGTCCGGGCTATGAGCTCATGCTGAGGCTTTATGGGAT 493
    |||||||
QY 573 cggcgtggaagacgagcgagcgagcgagtgagcgagacagaaactcaagagcgctt 632
    |||||||
Db 494 CCAACTTGAAGACCGGAACAGGCGCGTATGCGGTGACGAAGACTCCACACGCTT 553
    |||||||
QY 633 ccagaaactgaactgtgcgagccagaacaaactcgcgtatcaagcaatcctcaagtcgc 692
    |||||||
Db 554 ACACAACTCAAAAGCCAGACACAACTGCTGTTATACAGCATCTCAAGTCTGCT 613
    |||||||
QY 693 gtgtgagctgag 704
    |||||||
```

Db 614 GGCGAGCTGGG 625

RESULT 14
LOCUS BE391975 549 bp mRNA EST 21-JUL-2000
DEFINITION 601285002F1 NIH_MGC_44 Homo sapiens CDNA clone IMAGE:3606840 5',
mRNA sequence.
ACCESSION BE391975
VERSION BE391975.1 GI:9337340
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 549)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubln Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LCM257 row: c column: 01
High quality sequence start: 77
High quality sequence stop: 548.
Location/Qualifiers

FEATURES
Source
1..549
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3606840"
/tissue="uterus; placenta; fetal brain 00004"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCR7; Site: 1: XhoI; Site: 2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G) library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 126 a 133 c 184 g 106 t

ORIGIN

Query Match 24.9%; Score 398; DB 35; Length 549;
Best Local Similarity 97.7%; Pred. No. 3.7e-72;
Matches 467; Conservative 0; Mismatches 5; Indels 6; Gaps 6;

QY 61 tggagagagagagagagagatgagagagac-gcgagagagagagagagagagagc-gcg 118
DB 72 tggtagagagagagagagagatgagagagac-gcgagagagagagagagagagagc 131
QY 119 aggc 178
DB 132 aggc 191
QY 179 gggcg 237
DB 192 gggcg 251
QY 238 acgag 295
DB 252 acgag 311
QY 296 atgggag 354

Db 312 ATGGGAGACGCCAACCAGTGTCTTACAGAAATGATCCGCTTCCTCCAGCGC 371

QY 355 tggtag 414
DB 372 tggtag 431
QY 415 cactctacacacagtgctgttccctcgcgagagagagagagagagagagagagagagag 474
DB 432 cactctacacacagtgctgttccctcgcgagagagagagagagagagagagagagagag 491
QY 475 cccctcagcctcag 532
DB 492 CCCCTCAGCTCAGGAGAGTGGAGTGTAAAGCTCCAGAGATCCAGAGCGCG 549

RESULT 15
LOCUS A1816230 400 bp mRNA EST 09-JUL-1999
DEFINITION au45h07.Y1 Schneider fetal brain 00004 Homo sapiens CDNA clone
IMAGE:2517757 5' similar to contains element MSRI repetitive
element; mRNA sequence.
ACCESSION A1816230
VERSION A1816230.1 GI:5431776
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 400)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
White, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B.,
WashU-NCI human EST Project
Unpublished (1997)
Other ESTs: au45h07.x1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40NP from Glbco
High quality sequence stop: 383.
Location/Qualifiers

FEATURES
Source
1..400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2517757"
/tissue="fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site: 1: SstI; Site: 2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAG-3'
and 3' adaptor sequence:
5'-GAG-3'.
The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."

BASE COUNT 98 a 109 c 138 g 55 t

ORIGIN

Query Match 24.5%; Score 392.2; DB 13; Length 400;
 Best Local Similarity 99.2%; Pred. No. 5,6e-71;
 Matches 394; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 27 gccgagcatggagaccccgactgcactccacctgggaagagagagagatgcga 86
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 4 gccgagcatggagaccccgactgcactccacctgggaagagagagatgcga 63
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 87 ggaacgagagagagagagactgcgaagagagagagagagagagagagagag 146
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 64 GGACCGGAGAGACGAGAGACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 147 agggagagagagagagagagagagagagagagagagagagagagagagag 206
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 124 AGGGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 183
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 207 cagaatgacagaggtccagaactgagcagcagcagcagcagcagcagcagc 266
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 184 CAGAAATGACAGAGGTCCAGAACTGGCGAGCCAGAGAGAGAGAGAGAGAGAG 243
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 267 ctatccggaactggtggaaagagactgcaatgggaaacgccaactgattctac 326
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 244 CTATCCGATCTGTGTGAACGAGACTGCAATGGGGACACGCCAACTGAGTTT 303
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 327 aaatgagatccgcttcctcccaagcgctgtttcatgtgagacattctcaga 386
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 304 AAATATGATCCGCTTCTCTCCACGCGTGTTCATTGAGACATTTCTCAGAA 363
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 387 ggacaactatgacctctcttgagagacaatcactctac 423
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 364 AGCAACTATGACCTCTGTGAGACAACTACTCTCTAC 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
    
```

Search completed: October 12, 2000, 12:02:56
 Job time: 24164 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 18:05:10 ; Search time 563.62 Seconds
(without alignments)
1564.983 Million cell updates/sec

Title: US-09-431-843B-9

Perfect score: 2348
Sequence: 1 tagaattcagcgccgcctga.....cccttcgtgagtgaggggcg 2348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36: *
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT: *
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT: *
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT: *
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT: *
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT: *
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT: *
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT: *
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT: *
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT: *
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT: *
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT: *
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT: *
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT: *
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT: *
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT: *
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT: *
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT: *
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT: *
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT: *
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT: *
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	684.2	29.1	789	20	X40044
2	545	23.2	802	20	X40045
3	300.4	12.8	375	20	V90541
4	87.6	3.7	114955	20	X53491
5	84.4	3.6	114955	20	X53491
6	75.8	3.2	789	20	X40044
7	75.2	3.2	10732	21	A10594
8	73.6	3.1	1542	10	N90025
9	71.2	3.0	1925	20	X90924
10	70.2	3.0	35100	20	V73802
11	70.2	3.0	137507	19	V19941
12	67.2	2.9	799	19	V55831

C 13	67.2	2.9	5452	20	X90923	Anti-sense strand
C 14	67.2	2.9	8705	20	Z23778	Vector pshuttle DN
15	67.2	2.9	9600	19	V21683	Vector plasmid pck
16	67.2	2.9	10380	20	Z22248	Nucleotide sequenc
17	67.2	2.9	10596	14	O51731	Plasmid pCISBON f
18	67.2	2.9	10596	17	T40348	Plasmid pCISBON f
19	67.2	2.9	10596	20	X15650	Nucleotide sequenc
20	64	2.7	795	19	V55830	Flga insert stabl
21	62.6	2.7	390	13	O21833	Randomising Oligon
22	62.6	2.7	390	14	O36859	PCR primer for 5'
23	61.8	2.6	8438	15	O73500	DNA encoding Pseud
C 24	60.2	2.6	1337	20	Z17263	Human gene express
C 25	60.2	2.6	32207	20	V73805	KSHV LTR DNA (nucl
C 26	60.2	2.6	137507	19	V19941	KSHV long unique c
C 27	60	2.6	117213	19	V62176	HSV-2 strain S85 C
C 28	59.2	2.5	600	6	N50853	Sequence encoding
C 29	58.4	2.5	150	20	V64956	Mouse histone H2B
C 30	58.4	2.5	51259	18	X83007	Partial mouse WRN
31	58	2.5	1218	21	A02488	Human colon cancer
32	57.4	2.4	1593	21	A02504	Human colon cancer
C 33	57.2	2.4	309	10	N90579	PDF9.3 CDNA insert
34	56.2	2.4	2277	19	V13836	Human telomerase p
35	56.2	2.4	2277	19	V05372	Human colon cancer
C 36	56	2.4	1000	21	A02484	Partial mouse WRN
C 37	56	2.4	16442	18	X83006	Human prostate pro
38	54.8	2.3	768	18	T84941	CDNA encoding a pr
39	54.8	2.3	768	20	X35871	Leishmania donovan
C 40	54.8	2.3	1091	16	O85844	CENP-B CDNA. Homo
41	54.2	2.3	2643	14	O39212	Human lung tumour
42	54	2.3	2109	20	Z07194	Leishmania donovan
C 43	53.2	2.3	761	16	O89794	Human enzyme-relat
C 44	53.2	2.3	6225	20	X55273	Human adenosine re
C 45	53.2	2.3	6225	21	A34721	

ALIGNMENTS

RESULT 1	
X40044	X40044 standard; DNA; 789 BP.
XX	
AC	X40044;
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Prostate cancer associated gene.
XX	
KW	Cancer associated antigen; diagnosis: research; treatment: human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	W09904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998; 98WO-US14679.
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0886164.
PR	10-OCT-1997; 97US-0061599.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI	Pfreundschuh M, Sahlin U, Scanlan MJ, Stockert E;
PI	Tureci O;
XX	

DR WPT; 1999-132448/11.

PT New Isolated cancer associated nucleic acids and polypeptides -
PT Isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers

PS Claim 67; page 646; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

SD Sequence 789 BP; 157 A; 259 C; 275 G; 93 T; 5 other;

Query Match	29.18;	Score 684.2;	DB 20;	Length 789;
Best Local Similarity	93.98;	Pred. No. 4.5e-115;		
Matches 742; Conservative	0;	Mismatches 43;	Indels 5;	Gaps 3

[illegible]

Accession	Sequence	Position
OY 2071	aaggaag-tgccctcggcgagctcttgagccctccctctcgaaggctgagctccg	212-23
Db 661	aaggaagctggcccgctggcgctctctgcccctcgcctctctgaaggctctggmctccg	720
OY 2130	gaactgctgcggactccctcgaagctctctcttcggagccgtgacccatgacccacaagt	218-39
Db 721	gaactmctggagctctcccttaaggtcttggtt--gtgacccctgacccatgancaccacaatg	778
OY 2190	ctggcctcct 2199	
Db 779	ctgagctct 788	

RESULT	2
X40045/c	
ID	X40045 standard; DNA; 802 BP.

DT 02-JUL-1999 (first entry)

DE Prostate cancer associated gene.

KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.

OS Homo sapiens.

PN W09904265-A2.

PD 28-JAN-1999.

15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR	10-OCT-1997:	97US-0061599
PR	17-OCT-1997:	97US-0896104

PR	10-OCT-1997;	97US-0061165.
PR	10-OCT-1997:	97US-0948705.

PK 11-UCJ-1997; 9/GB-0021697.

PA (LUDWIG) LUDWIG INST CANCER H
XX

PI Chen Y, Gout I, Gure A, C
PI pfreundschub M, sabia n
PI

PI tureci O;
XX

DR WPI; 1999-132448/11.
YY

PT New isolated cancer associated

PT for the diagnosis, monitoring

PS Claim 67; Page 646; 787pp; E

CC The invention relates to a method

by expression of a human cancer associated antigen precursor coded for by a nucleic acid molecule (NAM). The method comprises: (a) contacting a biological sample isolated from a subject with an agent that specifically binds to the NAM, an expression product or a fragment of an expression product complexed with an HLA molecule; and (b) determining the interaction between the agent and the NAM or the expression product as a determination of the disorder. The products and methods can be used in the diagnosis, monitoring, research, or treatment of conditions characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer.

Sequence 802 BP; 112 A; 248 C; 277 G; 158 T; 7 other;

QY	1891	gagaccccaaggcccccccggagaagcctgcaggggaagaacacagcgagagcccatcg	1950
Db	104237	GCGGNNHNNNSGCCGGCCGCGGNHHNNSNGCCGCGCCGCAGNNHNNSNGCCGGCCGACNN	104178
QY	1951	gagaccaccaagccccagccccgcagcagc	1979
Db	104177	HNNNSGCCGCGCCGACGAGNNHNNNSGCC	104149
RESULT	5		
ID	X53491	standard; DNA; 114955 BP.	
XX	AC	X53491;	
XX	DT	05-JUL-1999 (first entry)	
DE	XX	Human adenosine A1 receptor antisense oligonucleotide fragment.	
KW	KW	Antisense oligonucleotide; multiple target; antisense treatment;	
KW	KW	impaired respiration; inflammation; lung disease;	
KW	KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;	
KW	KW	acute asthma; allergy; asthma; impeded respiration;	
KW	KW	respiratory distress syndrome; pain; cystic fibrosis;	
KW	KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;	
KW	KW	chronic obstructive pulmonary disease; leukemia; lymphoma;	
KW	KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;	
KW	KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;	
KW	KW	prostate cancer; ss.	
OS	OS	Synthetic.	
PX	PX	MO9913886-A1.	
PN	PN		
PD	PD	25-MAR-1999.	
XX	XX		
PF	PF	17-SEP-1998; 98WO-US19419.	
XX	XX		
PR	PR	09-JUN-1998; 98US-0093972.	
PR	PR	17-SEP-1997; 97US-0059160.	
PA	PA	(UYEC-) UNIV EAST CAROLINA.	
PI	PI	Nyce JW;	
PT	PT	WPI; 1999-229400/19.	
PS	PS	New antisense oligonucleotides used in treatment of, e.g. pulmonary vasoconstriction	
XX	XX	Disclosure; Page 37; 120pp; English.	
XX	XX	The specification describes antisense oligonucleotides (X52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences X55272-74. These multiple target oligonucleotides (specifically X55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.	

[illegible]

[illegible]

PR	10-OCT-1997;	97US-0061765.
PR	10-OCT-1997;	97US-0948705.
XX	11-OCT-1997;	97GB-0021697.
PA	(LUDMW-) LUDWIG INST CANCER RES.	
XX		
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;	
PI	Pfeunnschuh M, Sahin U, Scanlan MJ, Stockert E;	
PI	Tureci O;	
XX		
DR	WPI; 1999-132448/11.	
XX		
PT	New isolated cancer associated nucleic acids and polypeptides -	
PT	isolated using sera from cancer patients, used to develop products	
XX	for the diagnosis, monitoring or treatment of cancers	
XX		
PS	Claim 67; Page 646; 787pp; English.	
CC		
CC	The invention relates to a method for diagnosing a disorder characterised	
CC	by expression of a human cancer associated antigen precursor coded for by	
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a	
CC	biological sample isolated from a subject with an agent that specifically	
CC	binds to the NAM, an expression product or a fragment of an expression	
CC	product complexed with an HLA molecule; and (b) determining the	
CC	interaction between the agent and the NAM or the expression product as a	
CC	determination of the disorder. The products and methods can be used in	
CC	the diagnosis, monitoring, research, or treatment of conditions	
CC	characterised by the expression of various cancer associated antigens.	
CC	The invention provides nucleic acid sequences and encoded polypeptides	
CC	which are cancer associated antigen precursors expressed in human breast	
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and	
XX	lung cancer.	
XX		
SQ	Sequence 789 BP; 157 A; 259 C; 275 G; 93 T; 5 other;	
	Query Match	3.2%; Score 75.8; DB 20; Length 789;
	Best Local Similarity	49.4%; Pred. No. 1.4e-05;
	Matches 197; Conservative	0; Mismatches 202; Indels 0; Gaps 0
QY	1598 cagggcccaaggcagcagaactctgacagggagacgcagccagcgagaccatcgagacc	1657
DB	564 CTGGCTATCCCTTGTAAGGTCGCCGGCGCTGGGGCTCGGGGTCCGATGGGCTTCGG	525
QY	1658 caggcccccccagcaggaactctgacagggagagagccagccagcgagaccatcgagacc	1717
DB	524 CTGGCTGTGTCCTGCAGAGTCTCTCGGGGGCGGGGCTTCGAGTGGGCTTCGG	465
QY	1718 caggctctgcgccagcaggaactctgacagggagagacgcagccagaccatcgagacc	1777
DB	464 CTGGCTGTGTCCTGCAGAGTCTCTCGGGGGCGGGGCTTCGAGTGGGCTTCGG	405
QY	1778 caggcccccaaggcaggaactaaagaagtatgaaaccaaagcccgagaccatcgagacc	1837
DB	404 CTGGCTATCCCTTGTAAGTCTGCGGGCTGGGGCTTCGAGTGGGCTTCGG	345
QY	1838 caggcccccccagcaggaactctgacagggagagacgcagccagcgagaccatcgagacc	1897
DB	344 CTGGCTGTGTCCTGCAGAGTCTCTCGGGGGCGGGGCTTCGAGTGGGCTTCGG	285
QY	1898 caggcccccccagcaggaactctgacagggagagacgcagccagcgagaccatcgagacc	1957
DB	284 CTGGCTGTGTCCTGCAGAGTCTCTCGGGGGCGGGGCTTCGAGTGGGCTTCGG	225
QY	1958 caggcccccccagcaggaacttaagaagtatgagccag	1996
DB	224 CTGGCTGTGTCCTGCAGAGTCTCTCGGGGGCGGGCTTCG	186
RESULT	7	
ID	A10594	
ID	A10594 standard; DNA; 10732 BP.	

DG 504 ggcagagcagagagagagcagagcagagagagagcagagagagcagagagcagag 563
QY 1193 tggagctgagagccgagagagcagagccacagagccagagctgctccagag 1252
DB 564 aggggacagagagagagagagagagcagagagagagagagagagagagagag 623
QY 1253 tggagaaatcgctgctgatttggaggggtgtgtccctcagcagcagcagcag 1312
DB 624 aggggacag 683
QY 1313 ggaaccag 1372
DB 684 aggggacag 743
QY 1373 tggagagcag 1432
DB 744 aggggacag 803
QY 1433 acag 1436
DB 804 gcaag 807
RESULT 10
V73802
ID V73802 standard; DNA: 35100 BP.
XX
AC V73802;
XX
DT 25-FEB-1999 (first entry)
XX
DE KSHV LUR DNA (nucleotides 1-35,100).
XX
KM Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KM dihydrofolate reductase; LUR: long unique region; vaccine; prophylaxis;
KM diagnosis; treatment; HHV8; complement binding protein; v-CBP; SSBP;
KM ssDNA binding protein; transport protein; glycoprotein B; pol; vIL-6;
KM DNA polymerase; viral interleukin-6; BHV4-IE1 I; thymidylate synthase;
KM vMIP-II; BHV4-IE1 II; vMIP-I; capsid protein I; tegument protein I; ds.
XX
OS Kaposi's sarcoma-associated herpesvirus.
XX
PN US849564-A.
XX
PD 15-DEC-1998.
XX
PF 29-NOV-1996; 96US-0770379.
XX
PR 29-NOV-1996; 96US-0770379.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JU;
XX
DR WPI: 1999-069741/06.
XX
PT Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT dihydrofolate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
XX
PS Disclosure: Column 67-96: 109pp; English.
XX
CC This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for K1, ORF4 which encodes the complement binding protein
CC v-CBP, ORF6 which encodes a ssDNA binding protein (SSBP), ORF7 which
CC encodes a transport protein, ORF8 which encodes glycoprotein B, ORF9
CC which encodes DNA polymerase (pol), ORF10, ORF11, K2 which encodes viral
CC interleukin-6 (vIL-6), ORF02 which encodes dihydrofolate reductase
CC (DHFR), K3 which encodes BHV4-IE1 I, ORF70 which encodes Thymidylate
CC synthase, K4 which encodes vMIP-II, K5 which encodes BHV4-IE1 II, K6
CC which encodes vMIP-I, K7, ORF16 which encodes Bcl-2, ORF17 which encodes
CC capsid protein I, ORF18 and ORF19 which encodes Tegument protein I.

CC KSHV is a new human Herpesvirus (HHV8) believed to cause Kaposi's sarcoma
CC (KS) which is the most common form of neoplasm occurring in persons with
CC acquired immune deficiency syndrome (AIDS). The DHFR protein is useful
CC for vaccination, prophylaxis, diagnosis and treatment of a subject with
CC Kaposi's sarcoma and for detecting expression of a DNA virus associated
CC with Kaposi's sarcoma in a cell.
XX
SQ Sequence 35100 BP; 8703 A; 9395 C; 8921 G; 8081 T; 0 other:
Query Match 3.0%; Score 70.2; DB 20; Length 35100;
Best Local Similarity 48.8%; Pred. No. 0.00019;
Matches 221; Conservative 0; Mismatches 228; Indels 4; Gaps 1;
QY 1587 atcggaagaccagagccagccagcagagagagagagagagagagagagagag 1646
DB 24281 aacgagagagcccgagcagacccagagagcccgagcagcagcagagagcccgagca 24340
QY 1647 atcggaagaccagagcccgagcagagagagagagagagagagagagagagag 1706
DB 24341 cccagagagcccgagcagacccagagagcccgagcagacccagagagcccgagca 24400
QY 1707 atcggaagaccagagcccgagcagagagagagagagagagagagagagagag 1766
DB 24401 cccagagagcccgagcagacccagagagcccgagcagacccagagagcccgagca 24460
QY 1767 atcggaagaccagagcccgagcagagagagagagagagagagagagagagag 1826
DB 24461 cccagagagcccgagcagacccagagagcccgagcagacccagagagcccgagca 24520
QY 1827 atcggaagaccagagcccgagcagagagagagagagagagagagagagagag 1886
DB 24521 cccagagagcccgagcagacccagagagcccgagcagacccagagagcccgagca 24580
QY 1887 atcggaagaccagagcccgagcagagagagagagagagagagagagagagag 1946
DB 24581 cccagagagcccgagcagacccagagagcccgagcagacccagagagcccgag 24640
QY 1947 atcggaagaccagagcccgagcagagagagagagagagagagagagagagag 2006
DB 24641 cccgagcagcgc---accctcccgagagggagatcccgagcagacccctcccgag 24696
QY 2007 ggaagcagcagaggttgagagagcagagagtgga 2039
DB 24697 ggaaccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 24729
RESULT 11
V19941
ID V19941 standard; DNA: 137507 BP.
XX
AC V19941;
XX
DT 03-AUG-1998 (first entry)
XX
DE KSHV long unique coding region and terminal repeat.
XX
KM KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KM interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KM complement-binding protein; glycoprotein; capsid protein IV; infection;
KM immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KM lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KM HIV immune status; anti-inflammatory agent; therapy; ds.
XX
OS Kaposi's sarcoma-associated herpes virus.
XX
FH Key Location/Qualifiers
FT 1142..2794
FT CDS
FT /*tag= a
FT /product= complement-binding protein
FT CDS 8699..11236
FT /*tag= b
FT /product= glycoprotein B

XX	Epstein-barr virus.
OS	
XX	W09822577-A1.
XX	
XX	28-MAY-1998.
PD	
XX	
PF	17-NOV-1997; 97WO-IB01508.
XX	
XX	25-JUN-1997; 97US-0048945.
PR	15-NOV-1996; 96US-0030986.
XX	
PA	(MASU/) MASUCCI M G.
XX	
PI	Masucci MG;
XX	
DR	WPI; 1998-312463/27.
XX	
PT	New fusion proteins resistant to proteolytic degradation -
PT	comprising a core protein with a stabilising polypeptide comprising
XX	a peptide sequence containing glycine repeats
XX	
PS	Disclosure; Fig 4B; 120pp; English.
XX	
CC	This is a nucleotide sequence of the stabilising sequence-encoding
CC	insert. The invention provides a method for increasing the resistance
CC	of a core protein to proteolytic degradation that comprises linking or
CC	inserting onto or into the core protein a stabilising polypeptide of
CC	formula [(G1Yb)X(G1Yb)Y(G1Yc)Z]n where G1ya, G1yb, G1yc are 1-6
CC	sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC	Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC	be identical from n repeat to n repeat. Alternatively a nucleic acid
CC	encoding the stabilising polypeptide can be linked onto or inserted into
CC	a nucleic acid encoding a core protein. The fusion proteins of the
CC	invention are more resistant to degradation by proteases and, thus, have
CC	a longer half-life than the unfused core protein. The products can be
CC	used for treating autoimmune diseases, cancer and inflammation. In
CC	particular, the core protein may be an IkappaB regulator protein for the
CC	treatment of inflammatory bowel disease, or a nitroreductase protein
CC	which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC	or other pathological conditions. The fusion proteins can also be used in
CC	diagnostic methods such as in vivo imaging.
XX	
CC	
SQ	Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;
	Query Match 2.9%; Score 67.2; DB 19; Length 799;
	Best Local Similarity 47.7%; Pred. No. 0.00049;
	Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1.
QY	956 ggaaggtgtgagaggaagaagcccccggagaccacccagcagagccaccacaggtgc 1015
DB	82 ggcacagagagcagagagagggcgagagggcgacagagagggcgagggcgagggag 141
QY	1016 ggaactgtggcgacagagcctaacaggtgtg---ggcgaggtgtgacgagggccacgc 1072
DB	142 agggcgacagagcagagagggcgagagggcgagagggcgagggcgagggcgaggggc 201
QY	1073 cccgcagacgttgagagcccccgatgctggcgaccccttgagagggcgaggggagtgagcag 1132
DB	202 aggaagcagagagagagggcgagagggcgacagagcagagagagagggcgagagag 261
QY	1133 gggcgccagcagggagaaataagccggagcccttaagccccaagaagagcagaagagagc 1192
DB	262 ggcacagagagcagagagagggcgagagcagagagagagagggcgagggcgagggcgagag 321
QY	1193 tggagcgtgagccggcgagggcagccgccacacagagcagagccctcagagtgtcctcagagag 1252
DB	322 agggcgagagagagggcgagggcgagggcgagagcagagagagggcgagggcgagggcgag 381
QY	1253 tggagagaagtcctcgtgaatttggagaggtgtgtgccttcagcgacgggcaactcagagcgg 1312
DB	382 agggcgagagagggcgagggcgagggcgagggcgagagcagagagagagagggcgagggcgag 441

Oy	1313	ggaccaggaagtggcggtcaggacccctgaggagcgatgcagaccttgcgccaaacccc	1372
Dd	442	aggggcaggagcaggaggggcagagagcaggagggccaggagcaggagcaggagcagg	501
Oy	1373	tgggagccaaggttcgcccagacaagttgaggagaacggaagagtgtagtgaaggtc	1432
Dd	502	agggcagagaggggcagagcagagaggggcaggaggggcagagacagagagggcag	561
RESULT	13		
ID	X90923/c		
XX	X90923 standard; DNA; 5452 BP.		
AC	X90923:		
DT	17-JAN-2000 (first entry)		
DE	Anti-sense strand of PCWENBNA plasmid.		
XX			
KM	Anti-sense strand; Plasmid PCWENBNA; EBNA 1; episome; transfection;		
KM	Epstein Barr Virus Nuclear Antigen 1; origin of replication;		
KM	EBV oriP; eucaryotic host cell; recombinant cell line; ion channel;		
KM	multiple gene expression; receptor; transporter protein; gene therapy;		
KM	transcription factor; adhesion molecule; antisense therapy;		
XX	gene amplification; cell immortalisation; ss.		
OS	Epstein-barr virus.		
OS	Cytomegalovirus.		
OS	Synthetic.		
Key	Location/Qualifiers		
FT	CDS complement (3032..4957)		
FT	/*tag - a		
FT	/product= "EBNA 1 protein"		
XX	/note= "Epstein Barr Virus Nuclear Antigen 1"		
PX	WO9947647-A1.		
XX	23-SEP-1999.		
XX	12-FEB-1999; 99WO-US03307.		
PX	18-MAR-1998; 98US-0040961.		
PR	06-AUG-1998; 98US-0130114.		
PA	(PHAR-) PHARMACOPEIA INC.		
PI	DamaJ BB, Horlick RA, Robbins AK;		
DR	WPI: 1999-610610/52.		
PT	New method for expressing genes from recombinant eukaryotic cells,		
PT	useful for gene therapy -		
PS	Example 1; Fig 1; 86pp; English.		
XX			
CC	The present sequence is an anti-sense strand of commercially available		
CC	plasmid PCWENBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA		
CC	1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain		
CC	episomes containing EBV origin of replication (oriP) and a gene encoding		
CC	protein or RNA of interest. Eucaryotic host cells expressing EBNA 1		
CC	protein are transfected with these episomes to produce recombinant		
CC	cell lines expressing multiple genes of interest. This provides a		
CC	rapid and reliable method of stably expressing multiple genes in		
CC	transfected cells. The episomes are useful in the transfection of genes		
CC	encoding receptors, transporter proteins, ion channels, adhesion		
CC	molecules and transcription factors. The episomes carrying desired genes		
CC	can also be used to transfect cells in gene therapy, antisense therapy,		
CC	for gene amplification, cell immortalisation, etc.		
Sequence	5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;		


```

FT      /*tag- b
FT      /note- "EBV origin of replication"
FT      repeat_unit
FT      4928..5104
FT      /*tag- c
FT      /rpl_type- INVERTED
FT      /note- "AAV Inverted terminal repeat"
FT      repeat_unit
FT      7189..7355
FT      /*tag- d
FT      /rpl_type- INVERTED
FT      /note- "AAV Inverted terminal repeat"
FT      promoter
FT      5112..6734
FT      /*tag- e
FT      /note- "CMW Immediate-early enhancer/promoter"
FT      terminator
FT      6818..7050
FT      /*tag- f
FT      /note- "bovine growth hormone polyA sequence"
XX
XX      WO9806437-A2.
XX      19-FEB-1998.
XX
XX      13-AUG-1997; 97WO-US14465.
XX
XX      13-AUG-1996; 96US-0023867.
XX
XX      (CHIR ) CHIRON CORP.
XX
XX      Cohen F, Dubois-Stringfellow N, Darkl V, Innis MA;
XX      Murphy JE, Tetsuo U, Zukermann R;
XX      WPI; 1998-159296/14.
XX
XX      Polycationic agents based on alpha-amino acids, able to complex
XX      with nucleic acid - to facilitate its entry into cell, condense it
XX      and protect it against serum degradation, particularly for use in
XX      gene therapy
XX
XX      PS      Disclosure; Page 77-80; 100pp; English.
XX
XX      CC      This polynucleotide comprises the DNA sequence of vector plasmid
XX      PCWKM1R-EPI, which contains an Epstein-Barr virus (EBV) origin
XX      of replication from plasmid pCEP4, a coding region for EBV nuclear
XX      antigen A from pCEP4, a pair of inverted terminal repeats from
XX      adeno-associated virus, a cytomegalovirus enhancer/promoter, a
XX      bovine growth hormone polyA sequence, and a kanamycin resistance
XX      selectable marker. Polynucleotides encoding polypeptides, such as
XX      erythropoietin or leptin, and ribozymes and antisense
XX      polynucleotides can be inserted into the vector. The vector is
XX      preferred for use in novel compositions and methods for improved
XX      polynucleotide delivery into cells. In these methods, polycationic
XX      agents are used to increase the frequency of uptake of a
XX      nucleic acid (see also V21684-86) into a cell. The polycationic
XX      agent can condense with the nucleic acid and inhibit serum and/or
XX      nuclease degradation of the nucleic acid. The nucleic acid can be
XX      a vector, may express a therapeutic protein or a vaccinating viral
XX      or cancer antigen, or is itself therapeutic (antisense or
XX      ribozyme). The methods and compositions can be used in the gene
XX      therapy of many diseases.
XX
XX      SQ      Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

```

```

Query Match      2.9%; Score 67.2; DB 19; Length 9600;
Best Local Similarity 47.7%; Pred. NO. 0.0006;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

```

```

QY      956 ggaaggtgaggaaggaagcccccggggagcccgaccagagagccagcccgaggtc 1015
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      753 ggcagagcagagagagagagagagagagagagagagagagagagagagagag 812
QY      1016 ggaacctgtggccagagatagcaaggtg---ggcagaggttgacagagggcccccagc 1072
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      813 aggggcagagagagagagagagagagagagagagagagagagagagagagagaggggc 872

```

```

QY      1073 cagcagcgtgagaccagagatcgaggaccctctgagagagagccagagagatgagcag 1132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      873 agcagcagagagagagagagagagagagagagagagagagagagagagagagagag 932
QY      1133 gggccacaggggaagatagcccgagcccttaagccccaagagagcaagagagagagc 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      933 ggcagagcagagagagagagagagagagagagagagagagagagagagagagagagag 992
QY      1193 tggagctgagccgagcagagagcagcccgcccaagagagagagagagagagagagag 1252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      993 agggcagagagagagagagagagagagagagagagagagagagagagagagagagag 1052
QY      1253 tggagagatgcctctgagatttgagaggggtgtgcccctcagcagagagagagagag 1312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1053 agggcagagagagagagagagagagagagagagagagagagagagagagagagag 1112
QY      1313 ggaaccagagaggtggcgtcagagaccctggggagagcagtgcaagccctgcccgaacccc 1372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1113 agggcagagagagagagagagagagagagagagagagagagagagagagagagagag 1172
QY      1373 tggagagcagaggtggccgacagagtgaggaagcggagagagagagagagagagagag 1432
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1173 agggcagagagagagagagagagagagagagagagagagagagagagagagagagag 1232

```

Search completed: October 12, 2000, 18:22:31
Job time: 31809 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 12:02:56 ; Search time 7878.83 Seconds
(without alignments)
1842.562 Million cell updates/sec

Title: US-09-431-843B-9
Perfect score: 2348
Sequence: 1 tagaatcagcgccgctga.....ccctctgagtgagggggagc 2148

Scoring table: IDENTITY_NMC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_esthum21:*
64: em_esthum22:*
65: em_esthum23:*
66: em_esthum24:*
67: em_esthum25:*
68: em_esthum26:*
69: em_esthum27:*
70: em_esthum28:*
71: em_esthum29:*
72: em_esthum30:*
73: em_esthum31:*
74: em_esthum32:*
75: em_esthum33:*
76: em_esthum34:*
77: em_esthum35:*
78: em_esthum36:*
79: em_esthum37:*
80: em_esthum38:*
81: em_esthum39:*
82: em_esthum40:*
83: em_esthum41:*
84: em_esthum42:*
85: em_esthum43:*
86: em_esthum44:*
87: em_esthum45:*
88: gb_gsa1:*
89: gb_gsa2:*
90: gb_gsa3:*
91: gb_gsa4:*
92: em_gsa1:*
93: em_gsa2:*
94: em_gsa3:*
95: em_gsa4:*
96: gb_gsa15:*
97: gb_gsa16:*
98: gb_gsa17:*
99: gb_gsa18:*
100: gb_gsa19:*
101: em_gsa5:*
102: em_gsa6:*
103: em_gsa7:*
104: em_gsa8:*
105: em_gsa9:*
106: em_gsa10:*
107: em_gsa11:*
108: em_gsa12:*
109: em_gsa13:*
110: gb_gsa5:*
111: gb_gsa6:*
112: gb_gsa7:*
113: gb_gsa8:*
114: gb_gsa9:*
115: gb_gsa10:*
116: gb_gsa11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	636.6	27.1	729	34	BE264806 601192867
2	596.6	25.4	868	34	BE277227 601178592
3	592.6	25.2	751	35	BE387193 601277069
4	586	25.0	682	35	BE385629 601278145
5	530	22.6	546	21	AM246100 2821312.5
6	523	22.3	606	21	AM338856 2821312.5
7	514	21.9	522	14	AL120919 2821312.5
8	508.4	21.7	526	24	AM732857 601303311
9	506.2	21.6	591	35	BE410591 601303311
10	477.8	20.3	493	35	BE336852 601298683
11	474.2	20.2	551	34	BE207411 601298683
12	462.6	19.3	550	21	AM249563 2821312.3
13	453.6	19.3	529	35	BE398108 601290328
14	442	18.8	1286	35	BE409569 601298683
15	435.4	18.5	757	35	BE391750 601298683
16	426.6	18.2	526	22	AM513841 601298683
17	425.8	18.1	433	10	AI459808 601298683
18	424.8	18.1	502	34	BE206098 601298683
19	423	18.0	432	13	AI816189 601298683
20	415	17.7	432	21	AM250708 601298683
21	413.4	17.6	427	34	BE302499 601298683
22	400.8	17.1	654	34	BE282484 601103075
23	398	17.0	549	35	BE391975 601285002
24	392.2	16.7	400	13	AI816230 601298683
25	390.6	16.6	473	19	AM072475 601298683
26	376.4	16.0	454	13	AI926535 601298683
27	375.4	16.0	886	34	BE274933 601298683
28	370.2	15.8	445	11	AI590782 601298683
29	365.4	15.6	451	12	AI634247 601298683
30	359.4	15.3	436	10	AI432006 601298683
31	355.4	15.1	590	8	AI115047 601298683
32	354.8	15.1	735	36	BE534864 601298683
33	339.4	14.5	559	9	AI173869 601298683
34	330.8	14.1	559	23	AM631605 601298683
35	306.4	13.0	414	21	AM326593 19380 MAR
36	299.2	12.7	500	24	AM762644 601298683
37	297.8	12.7	496	35	BE482073 601298683
38	296.4	12.6	379	36	BE504832 601298683
39	291.2	12.4	506	21	AM321684 601298683
40	278.8	11.9	554	25	AM963575 601298683
41	272.8	11.6	1050	34	BE288040 601094726
42	259.8	11.1	447	13	AI786854 601298683
43	254.4	10.8	256	36	BE551638 601298683
44	252.8	10.7	355	7	AA865212 601298683
45	251	10.7	363	37	H31247 601298683

ALIGNMENTS

RESULT	1	BE264806
944	tcagaggtccaggaaggtgaggaaggaagcccccggagaccgacagagca	1003

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	421	TCGAGGCTCCAGGAAGGTGAGAGGAAGAGCCCGGGAGCCTCCGACACAGAGCCA	480					
Qy	1004	gcaccagaggtctcgagcctgtgtggcccaagcctagcaaggtgtggggcaggtgtgacg-ag	1062					
Db	481	GCACCCAGAGGTCGCGACCTGTGGCCAGAGCATACAGAGGGTGGGCGCAGAGGTGAGCAAG	540					
Qy	1063	ggggcccaagcagagaggtgtgagagcccaagatcgagacccctggagagagcagaggg	1122					
Db	541	GGGCCCCAGCCAGGAGGCTGTGAGCCCAAGATGCGGGAGCCCTGTGAGAGGAGCCA-GGG	599					
Qy	1123	gatgagcagaggggagcagcaggggaagatagcgcagagccttaagcccccaagagagcaag	1182					
Db	600	GATGAGCAGAGGGGCGCCAGGGGAGATAGGCCCGGAGCCCTTAAGGCCCAAGAGAGCAAG	659					
Qy	1183	aagaggaagcttgagagctgagcccgagcgggagcagcgcgccacag	1225					
Db	660	-AGAGCAAGCTGAGACTGCGCGGGGAGCAGCCCAAAAGCAG-701						
RESULT	2							
LOCUS	BE277227	868 bp	mRNA	EST	13-JUL-2000			
DEFINITION	601178592P1 NIH_MGC_20	Homo sapiens	CDNA	clone	IMAGE:3050950 5'			
ACCESSION	BE277227							
VERSION	BE277227.1	GI:9152194						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	EMBL:U00001, Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.							
AUTHORS	NIH-MGC	http://www.ncbi.nlm.nih.gov/MGC/						
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)							
JOURNAL	Unpublished (1999)							
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: ATCC/DCTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: Image.Lnl.gov Plate: LNC97 row: h column: 23 High quality sequence stop: 742.							
FEATURES	Location/Qualifiers							
SOURCE	1..868							
	/organism="Homo sapiens"							
	/db_xref="taxon:9606"							
	/clone="IMAGE:3050950"							
	/clone_lib="NIH_MGC_20"							
	/tissue_type="melanotic melanoma"							
	/lab_host="DH10B (phage-resistant)"							
	/note="Organ: skin; Vector: pOTR; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGGACGAGC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald W. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."							
BASE COUNT	171 a 245 c 294 g 157 t							
ORIGIN								
Query Match	25.4%; Score 596.6; DB 34; Length 868;							
Best Local Similarity	88.1%; Pred. No. 4.8e-111;							
Matches	760; Conservative 0; Mismatches 74; Indels 29; Gaps							
44	ccgactgcactcaactctgtgagagagcagagagatcgagagacgcggagagcag 103							

D	b	1	CCGACTGGACATCCACCCTGGGGAGGAGCAGAGAGATCGCGAGAGACCGGAGGACGAGG	60
Q	y	104	actgc-gaggaacgcgagagcc-----gccgcgcgaaggagcggagc	145
D	b	61	ACTGCTGAGAGACGGCGATGGCGCTGCCGTGCCTGAGTGAATGTGTGACTTCAGAGGTGAC	120
Q	y	146	caagggaacgagagacaggaagtcggaagagccgcggcgcgcgccccagctcgttccagt	205
D	b	121	TGAGGACGATGGATTTCCGATTGATGCTTGCTGGTGCGGCTCGGCGCCAGCTCGTTCCAGT	180
Q	y	206	ccaagaatcacagaggtgccagaaac-tgcggaagcccaagagacatgtgttaagtlalcggaac	264
D	b	181	CCAGATATGACAGAGGTCCAGAAACTGTGGGACCAACAGAGACAATGTGTAGGTATCGCGCAC	240
Q	y	265	aacttcgggatctgtytgaaacgagactgtcaa tggggacacgcaccaacttgattttcac	324
D	b	241	AACCTCCGGATCTTGTTGGAACGAGACTGCATATGGGGACACGCCAACAACCTGATTTCTAC	300
Q	y	325	agaatatgataatccgtttctcttgcccacagcgctgttcatatgtaggaatttttcagaactbg	384
D	b	301	ACAAAATGAGATCCGGTTCCTGCCCAACGGCGTTTCATTTAGGAGCATTTCTTCAAAACCTGG	360
Q	y	385	acggacaaciatgaacctctcttgagagacatacctcciaatcacaaagttgctgttcccttg	444
D	b	361	ACGGACAATATGACCTCTTGTAGAGACAATACTCTTAATCAATCCAGTGGCTTTTCTCTG	420
Q	y	445	cgaagaccaggaagtgtaacttggcatlbgcaagccctcaccgctcagaagagltcgaggtgtt	504
D	b	421	CGAGAACCAGGAGTGATGGATGGGATCCCAAGCCCCCTCACGCTCAGGAGAGGTGAGSTGYTT	480
Q	y	505	aaaagctcccaaggagatccagaagagcgcttgtccggcctaagctcatagtcttggtcttc	564
D	b	481	AAAACTCCCAAGAGATCCAGAGAGCGGCTTGTCCGGGCTTCAGAGCTCATATCTTGCGCTTC	540
Q	y	565	tac-ggagatccggctctggaagaccagagacagggcaacggttggccgaagacagaactacca	623
D	b	541	TACGGGGATCCGGCTCGGAGAGACCAAGAGCACGGGACAGGTGGGGCCGAGACAGAACTACCA	600
Q	y	624	gaagcgcttcagaacacttgaacttggcgagcgacacaacaacacttcgcatalcaacgcatacct	683
D	b	601	GAAAGCGTTCCAGAACCTGAACTGGCGCAGCCACACAACTCCGCAATCACACGCATTCCT	660
Q	y	684	caatgcgcgcgtgtgagcttgagcttcgagcaattccag--cgccactggtcgtcttttc	741
D	b	661	CAGATCGCTGGGTGAGCTGGGCTCTGACCACTTCAGGGCCGCTCGTGGCTCTTC	720
Q	y	742	ctgga-ggaagaagctggtgag--cggaagttgcggaggg--tgcygcaaggtgaccttg	796
D	b	721	CTGGGAGGAGAGCGCTGTGCGGGCGGTAGCTGCCGGGGGGGTGTCGGGAGAGGTGCCCTGG	780
Q	y	797	--acaacttcattgttcgcgtgtgcgttgcgcgaacacagcgccgcagcttgytgcaattcgc	854
D	b	781	GACTTACTTCAATGATCGGCGCTGTGCGTATGCGACACAGGCTTCGCAAGCTGTGGCCTTCGC	840
Q	y	855	ctggagacacattccggcccgct	877
D	b	841	CTTGGAGACCTTTCGGGCCCGT	863
<hr/>				
RESULT	3			
BE387193				
LOCUS	BE387193	751 bp	mRNA	EST
DEFINITION	601277069P1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3618136 5',			21-JUL-2000
ACCESSION	BE387193			
VERSION	BE387193.1	GI:9332558		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 751)			
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .			

QY 219 gtcagaactgagcagccacagaggaacatgtgtatgtatcgcaacatccgatct 278
 Db 183 gtcagaactgagcagccacagaggaacatgtgtatgtatcgcaacatccgatct 242
 QY 279 ggtggaacagagctcaatggggacacgccaactgagttctacagaatgatccg 338
 Db 243 ggtggaacagagctcaatggggacacgccaactgagttctacagaatgatccg 302
 QY 339 ctctctgccaacagctgtttctatgtagagactctctcagaactgagcgaactatga 398
 Db 303 ctctctgccaacagctgtttctatgtagagactctctcagaactgagcgaactatga 362
 QY 399 cctctctgagcaactcactcctacatccagtggtgttctctctcgaggaacagagat 458
 Db 363 cctctctgagcaactcactcctacatccagtggtgttctctctcgaggaacagagat 422
 QY 459 gaactggaatgcacacccctcagctcagagaggtcgaagtgtttaaagctcccaaga 518
 Db 423 gnaactggcatgccagccctcagctcagagaggtcgaagtgtttaaagctcccaaga 482
 QY 519 gat-ccaagagagcgc-tgtccgggctcagagctcagctgtgtgttctcagggatccg 576
 Db 483 gatccagagagcgcattgtccgggctcagagctcagctgtgtgttctcagggatccg 542
 QY 577 ctgagagacccgagcagagga-cagtgaggcagacagaaactacccaagcgttcca 635
 Db 543 ctgagagacccgagcagagga-cagtgaggcagacagaaactacccaagcgttcca 602
 QY 636 gaa--cctgaactgagcagcagcaacacacccctcagcatcacagatcctcaagtcgag 693
 Db 603 gaattcctgaactgagcagcagcaacacacccctcagcatcacagatcctcaagtcgag 662
 QY 694 tgtgagctgagcctcgagca 713
 Db 663 gtgacctgtggcctcgagca 682

RESULT 5
 LOCUS AM246100 546 bp mRNA EST 07-JAN-2000
 DEFINITION 2821312.5prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821312 5',
 mRNA sequence.
 ACCESSION AM246100
 VERSION AM246100.1 GI:6589093
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 546)
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Other ESTs: 2821312.3prime
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: DCTD/DRP cDNA Library Preparation: Ling
 Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
 Consortium (LINI). DNA Sequencing by: Berkeley MGC sequencing
 project clone distribution: MGC clone distribution information can
 be found through the I.M.A.G.E. Consortium/LINI at:
 www.bio.lini.gov/dbip/image/image.html Base Calling / Quality
 Scores: PHRED from University of Washington Genome Center. Vector
 Trimming: cross_match from University of Washington Genome Center
 PHRAP suite. Poly-T Identification: patchat.pl from Berkeley
 Drosophila Genome Project. Unification: patchat.pl from Berkeley
 http://www.genome.washington.edu
 plate: LIT66 row: H column: 17
 High quality sequence stop: 470.
 Location/Qualifiers
 1..546

FEATURES
source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2821312"
 /clone_lib="NIH_MGC_7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GCCACGAG(g). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 122 a 210 c 184 g 50 t
 ORIGIN

Query Match 22.6%; Score 530; DB 21; Length 546;
 Best Local Similarity 99.1%; Pred. No. 1.4e-97;
 Matches 533; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1459 gccacgaacttgcccttcgaggtcccttcgcccatcgaggacaccccaaggtcgagac 1518
 Db 9 gccacgaacttgcccttcgaggtcccttcgcccatcgaggacaccccaaggtcgagac 68
 QY 1519 agtgaacagcgggttgagaggaacacagaaggtcgacagcggcccaagaagtacccct 1578
 Db 69 agtgaacagcgggttgagaggaacacagaaggtcgacagcggcccaagaagtacccct 128
 QY 1579 gggagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1638
 Db 129 gggagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 188
 QY 1639 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1698
 Db 189 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 248
 QY 1699 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1788
 Db 249 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 308
 QY 1759 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1818
 Db 309 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 368
 QY 1819 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1878
 Db 369 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 428
 QY 1879 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1938
 Db 429 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 488
 QY 1939 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 1996
 Db 489 gagagcccatggagagaccacagcggcccaagcagagacttcaggggagagcagacc 546

RESULT 6
 LOCUS AM338856/c 606 bp mRNA EST 31-JAN-2000
 DEFINITION ha67f02.x1 NCI-GGAP_Pani Homo sapiens cDNA clone IMAGE:2878779 3',
 similar to TR:096029 096029 7-60.; contains element MSRI repetitive
 element ;, mRNA sequence.
 ACCESSION AM338856
 VERSION AM338856.1 GI:6835482
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	1 (bases 1 to 606)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL	Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Life Technologies catalog #: 1548-013 DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Seq primer: -40UP from Gibco High quality sequence stop: 421.
FEATURES	Location/Qualifiers
source	1..606 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2878779" /clone_1lb="NCI-CGAP_Pan1" /tissue_type="adenocarcinoma" /lab_host="DH10B" /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT	93 a 195 c 204 g 111 t 3 others
ORIGIN	
Query Match	22.3%; Score 523; DB 21; Length 606;
Best Local Similarity	97.5%; Pred. No. 3.8e-96;
Matches 540: Conservative	0; Mismatches 13; Indels 1; Gaps 1;
OY 1795	ggacctacaaggaatgagacgcagccgagagcccatcgagagccccaagcccgcccgccgca 1854.
Db 606	ggacctgcagaaggatgagccactgagaccatgcagacccccccagcccgccagca 547
OY 1855	ggacctgcagggagagccagccagccagccatctcgagagccccaagcccgcccgca 1914
Db 546	-gacctgcagggagagccagccacnagaccatcganagaccanagcccgcccgccgga 488
OY 1915	ggacctgcagggagagccagccagccatctcgagagccccaagcccgcccgca 1974
Db 487	ggacctgcagggagagccagccagccatctcgagagccccaagcccgcccgca 428
OY 1975	ggacctacaaggaatgagacgcagccagccagccagccatctcgagagcccgag 2034
Db 427	ggacctacaaggaatgagacgcagccagccagccagccagccatctcgagagcccgag 368
OY 2035	gtgaggtctctgcgaagtctggaagccctaaagaaagagtgccgctcggtctcttg 2094
Db 367	gtgaggtctctgcgaagtctggaagccctaaagaaagagtgccgctcggtctcttg 308
OY 2095	tcctcctctccctgtgtgcaagggctcgggggccctcggaagctgtgtggagccctcaggc 2154
Db 307	tcctcctctccctgtgtgcaagggctcgggggccctcggaagctgtgtggagccctcaggc 248
OY 2155	tcctgtcttgtagccgtgaccatgacccacagtgctgtgctctgtgtgggccaataag 2214
Db 247	tcctgtcttgtagccgtgaccatgacccacagtgctgtgctctgtgtgggccaataag 188
OY 2215	cagccacacgaagacgcgagagccctccaagggaaagcccaagccctcagagccctctggcc 2274
Db 187	cagccacacgaagacgcgagagccctccaagggaaagcccaagccctcagagccctctggcc 128
OY 2275	tggctgtgtctctccacccagctctcccccggcgccctgtcttggtaattgacccctc 2334
Db 127	tggctgtgtctctccacccagctctcccccggcgccctgtcttggtaattgaccccttc 68
OY 2335	tggagtgaggggggcg 2348

Db	67	TGAGATGGGGGGCGC	54
RESULT	7		
AL120919			
LOCUS	AL120919	522 bp	mrna
DEFINITION	DKFZP762B143..r1.762 (synonym: hmel2)		EST
ACCESSION	DKFZP762B143.5		25-FEB-2000
VERSION	AL120919		CDNA clone
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 522)		
TITLE	Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.		
JOURNAL	EST (Koehler, et al.)		
COMMENT	Unpublished (1999)		
	Contact: Koehler K		
	MIPS		
	Am Klopfersplitz 18a D-82152 Martinsried, Germany		
	This is the 5' sequence of the clone insert		
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;		
	Sequenced by BMR (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.		
	No sl sequence available.		
	This clone (DKFZP762B143) is available at the RZPD in Berlin.		
	Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.		
FEATURES	Location/Qualifiers		
source	1..522		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="DKFZP762B143"		
	/clone_lib="762 (synonym: hmel2)"		
	/russue_type="melanoma (Memo cell line)"		
	/dev_stage="adult"		
	/lab_host="DH10B"		
	/note="Vector: pSPort1; Site_1: NotI; Site_2: SalI"		
BASE COUNT	121 a 155 c 148 g 98 t		
ORIGIN			
Query Match	21.9%;	Score 514;	DB 14;
Best Local Similarity	99.0%;	Pred. No. 2.5e-94;	Length 522;
Matches 517;	Conservative 0;	Mismatches 5;	Indels 0;
			Gaps 0;
QY	221	ccagaactgcgagccacgaggaacatggttaagtatcgacaacatccgatatcg	280
Db	1	CCAGAAATTGCGAGCCACGAGGACATGTAGTATCGGCAACATTCGGATTCGG	60
QY	281	tggaaacgagactgtaatgaggagacacgccaacccgagttttacagaatatgacgcct	340
Db	61	TGGAACGAGACTGCAATGGGACACGCGCAACCGAGTTTCTAAGAAATAGTCCGCT	120
QY	341	tctgcgcccaagagctggtttacatctgagacatctcttcagaactgtagcgaactaac	400
Db	121	TCCTGCCCAAGGGCTGTTTCATTGAGGACATTTCTTCAGAACTGGACGACACATATGACC	180
QY	401	tcccttgaggaacatcactcctacatccagtgagctggtttcccttcgaggaaccagagtg	460
Db	181	TCCTTGAGGAAATCATTCTCTACATCCAGTGGCTGTTTCTCTCGGAACCGAGAGTGA	240
QY	461	actgagcatgccaagccctcaacgctcaaggagagtgctgagtggtttaaagaatccagaga	520
Db	241	ACTGGCATGCCAACCCCTCAACGCTCAGGGAGGTCGAGGGTGTAAAGCTCCACAGAGA	300
QY	521	tccagagagcgagctgttcgagggcctcagaagatattgctggagctctcaaggatccgctcg	580
Db	301	TCCAGAGACCGCTGTTCGGGCTACGAGCTCATGCTGGGCTCTCTACGGGATCCGGCTGG	360

QY	581	aggaccagagcgacgagcagcttgaggccagagacgactaccagaagcgcttcagagacc	640
Db	361	AGGACCGGAGGCACGGGCGCGGTGGGCGCAGACAGAACTACCAAGACGGCTTCCAGAAC	420
QY	641	tgaactggcgacgacacacacacacactccgcgcacacacgcatcctcaagtgcggtgagc	700
Db	421	TGAAGTGGCGGACGCCACACAACTCCGCATACACAGCATCTTAAGTCGCTGGGTAGC	480
QY	701	tgaagctcgagcacttcacagcgcaactggtgcgcttcctcc	742
Db	481	TGGCGCTCGAGACACTCCAGGCGCGGTGGTGGTCCGCTTCTTC	522
RESULT	8		
LOCUS	AW732857		
DEFINITION	bbl51g10.y1 NIH-MGC: 21 Homo sapiens cDNA clone IMAGE:2963010 5'		
ACCESSION	AW732857		
VERSION	AW732857.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 526)		
TITLE	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
CONTACT	Robert Strausberg, Ph.D.		

FEATURES	Source	Location/Qualifiers
1..526		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="IMAGE:2963010"		
/clone_lib="NH.MGC.21"		
/tissue_type="choriocarcinoma"		
/lab_host="DH10B (phage-resistant)"		
/note="Organ: Placenta; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G) size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	108 a	168 c
ORIGIN	193 g	57 t

	Query Match	21.78;	Score 508.4;	DB 24;	Length 526;
	Best Local Similarity	97.9%;	Pred. No. 3.4e-93;		
	Matches 515;	Conservative	0;	Mismatches 11;	Indels 0;
				Gaps	
QY	1216	ccgcccaagagccaggccctcagatgtgctcagagtgatgaagatgcctgaattg	1275		
Db	1	ccgccacattgccagagccctcagagtgctcagagagtgagacagatgccttgaaattg	60		
QY	1276	gaggggtgtgtgcccttcagccagggcagagctcagaacggggagccacaggaagtgggcgttag	1335		
Db	61	gaggggtgtgcccttcagccagggcagagctcagaacggggagccacaggaagtgggcgttag	120		

QY	1336	gacccctggggaagcagctgagacccctgcgcgcgaacccctcggagagccaaggtgtgacgcgaag	1395
Db	121	gaccctgggggaagcagctgacgacctgcgcgcgaacccctcggagagccaaggtgtgacgcgaag	180
QY	1396	gtcagagaacccgaaagaagtgtgactgaagtgtctggggaacatgtcgtgtgcagtgat	1455
Db	181	gtgagagaacccgaaagaagtgtgactgaagtgtctggggaacatgtcgtgtgcagtgat	240
QY	1456	ggtgcccaacacttggcccttgcggtccctcctgcacatcgggacaccccaagcttga	1515
Db	241	ggtgcccaacaccttgaccttgcgggtccctcctgcacatcgggacaccccaagccttga	300
QY	1516	cacagttgaaagagggtttgagagagacacagaaggtcgaacggggcccaagaagtctcc	1575
Db	301	cacagtgaaagagggtttgagagagacacatgaagttcgaaacggggcccaagaaggtctcc	360
QY	1576	cccttgagagcccactcgagagaccccaagcccccaagcaagagacctctcagggagcagcca	1635
Db	361	cccttgagagcccactcgagagaccccaagcccccaagcaagagacctctcagggagcagcca	420
QY	1636	gccgagagcccactcgagagaccccaagcccccaagcaagacacctctcagggagcagcca	1695
Db	421	gccgagagcccactcgagagaccccaagcccccaagcaagacacctctcagggagcagcca	480
QY	1696	gccgagagcccactcgagagaccccaagcccccaagcaagacacctctcagggagcagcca	1741
Db	481	gccgagagcccactcgagagaccccaagcccccaagcaagacacctctcagggagcagcca	526

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 591)	NIH-MGC	http://www.ncbi.nlm.nih.gov/MGC/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Ling Hong/Rubin Laboratory			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
	Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnhi.gov			
	Plate: L100337 row: f column: 03			
	High quality sequence start: 95			
	High quality sequence stop: 591.			
FEATURES				
source				
	1..591			
	location/Qualifiers			

```

/seq_name="XhoI10B"
/clone_image="3637634"
/clone_id="NH_MGC_21"
/tissue_type="Choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit

```


Db 430 CAAGCGGGGAGGACAGAGTTCAGAGCAGAGAGTGTCTTTCGCAAGTGTG 371
QY 2058 gaagccttaagaaagagatgcccgtgctgttccctgtcctgtcagagg 2117
Db 370 GAACCTTAAGAAAGAGATCCCGTGGGGCTTGTGCTCCCTGCTCCAGTGCAGAGG 311
QY 2118 ctgggggctcgggagctgctgtgagctccctcagagctgtcgtcgtgacccgtgacca 2177
Db 310 CTGGGGCTCCGGAGCTGCTGGGGCTCCCTCAGGCTCTGCTGAGCCCGGAGCCCA 251
QY 2178 tgaaccacagtgctgtgctcctgtgtgggacactaagacacacaaagccgagagcc 2237
Db 250 TGACCCACAGTGTGCTGCTCTGTGGGGCACTATAGACACACAGAACCCCGAGAGCC 191
QY 2238 ctcaaggaaagcccaaggctgcagaaagcctcgtgctgtgtgtcttccaccagc 2297
Db 190 CTCAGGGAAGCCCAAGGCTGCAGAGACCTCTGGCTGTGTCTCCACACCCAGC 131
QY 2298 tctccctgccccctgtctgttaattgacacctctgagatggggggg 2348
Db 130 TCTCCCTGCGCCCTCTTTTGTAAATTGACCTTCTGAGTGGGGGGCG 80

RESULT 13
BE398108 529 bp mRNA EST 21-JUL-2000
LOCUS 601299328F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3620767 5',
DEFINITION mRNA sequence.
ACCESSION BE398108
VERSION BE398108.1 GI:9343473
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 529)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LDCM293 row: g column: 08
High quality sequence start: 44
High quality sequence stop: 522.

FEATURES
source
1..529
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3620767"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 104 a 160 c 198 g 67 t
ORIGIN

Query Match 19.3%; Score 453.6; DB 35; Length 529;
Best Local Similarity 98.6%; Pred. No. 4.2e-82;

Matches 489; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 725 caatgtccgctcttccctgagagagcgtgtgcggcgaggctgcggggtgtcgcc 784
Db 35 CTGTATCCGCTTCTTCTCGAGAGACACTGTGTGCGGGAGCTGTCCGGGGTGGCCG 94
QY 785 aagatgccccctgactactcatglttcgcgtgtgcgtgcagacacacagccgacagctg 844
Db 95 AAGATGCCCTGTGACTTCACTTCACTTGTGCGCGTGGCTGCCAGACACAGCGCCGACGTGG 154
QY 845 tgcacttgccctggagagacttcggccccgctgcaagttcgtctgggggtcccaagaca 904
Db 155 TGCACCTGCGCTGGGAGGACACTTCCGCCCTCGTCAATTGCTGGGGGCCCAAGACA 214
QY 905 agctcgagaggttcaagccacagctctcgtgcacatcgcctgagaggtccagagaggtg 964
Db 215 ACTCGGGAGGTTCACAGCCACAGCTCTGTGCCCATCCCTGAGAGGCTCCAGGAAGTTGG 274
QY 965 agagaaagaaagcccgaggagcccgacacagagccagaccaggtctgacactgtg 1024
Db 275 AGGAGGAAGGAAGCCCGGGGACCCCGACACAGAGCCAGCACCA -GTTGGACCTGTG 333
QY 1025 ggcacagacatagcaaggtgtgggagaggtgtgagcgagggccacagccagagctgtg 1084
Db 334 GGCACAGACTRACCAAGGCTGAGGCA -GTTGACGAGAGGGCCACACGAGAGCTGG 392
QY 1085 agcccaagatgcggagacccctgtgagagagccagagggatgagagcaggggccaaggg 1144
Db 393 AGCCCAAGATGCGGAGACCTCTGGAGAGAGCA -GGGATGAGAGCAGGGGGCCACGGGG 451
QY 1145 aagatagcgcgagcgcttaagcccaagagaaagaaagaaagctgagctgagcc 1204
Db 452 AAGATAGCGCGGAGCCCTTAAGCCCAAGAGAGCAAGAGAACTGAGTGAAGCC 511
QY 1205 ggcggagagcagcgcc 1220
Db 512 GCGGAGACACCGCC 527

RESULT 14
BE409569 1296 bp mRNA EST 21-JUL-2000
LOCUS 60129983F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629948 5',
DEFINITION mRNA sequence.
ACCESSION BE409569
VERSION BE409569.1 GI:9346019
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
REFERENCE 1 (bases 1 to 1296)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LDCM17 row: e column: 21
High quality sequence start: 2
High quality sequence stop: 532.

FEATURES
source
1..1296
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629948"
/clone_lib="NIH_MGC_21"


```

/lissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/site_1="XhoI;
/site_2="EcoRI; cdna made by oligo-dt priming.
/notes="Organ: placenta; Vector: pOTB7;
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      327 a      416 c      389 g      104 t
ORIGIN

```

Query Match	18.8%	Score 442	DB 35	Length 1296
Best Local Similarity	84.6%	Pred. No. 9.8e-80		
Matches 668	Conservative 0	Mismatches 90	Indels 32	Gaps 14
QY 515	aggaaatccagagac-ggcttgcggggccctacagatcatctgttgcttctacggatc	573		
Db 11	AGGAGATCCAGAGCTAGCTGTCCGGGCTTCAGAGCTCATCTGTGGGCTTCACGGATTC	70		
QY 574	cggctcggagagacccgagcgcaagg-graaggttgggcccgaagacagaagaatctacgaagcgctt	632		
Db 71	CGGCTGGAGGAGCCAGGACGGCGACGCTGGGGCCGAGACAGAACTCAACCAAGCGCTT	130		
QY 633	ccaagaacttgaaactgtgcgaagcacaacaacctccgaatcaacagcatctcaagtgc	692		
Db 131	CCAGAACTTGAAACTGGCCAGCACAACAACTCCGCTACACCCATCTCAATTCCT	190		
QY 693	gtgtgaagctgagcctcgaagcacttcgaagcgagcacttgcgtctcttccttgagagagac	752		
Db 191	GGGTAGCTGGGCCCTCGAGCACTTCCAAGCGCGCTGGCCCTTCTCTTGAGAGAGAC	250		
QY 753	gctgtgtgagcgagagcgctgcgggggtgtcggagagtgccctgaacttactatcttc-g	811		
Db 251	GCTGTGTGGGGGGAGCTGTCCGGGGGTGGGAGTGTGCTTGAGACTTATATTTCAG	310		
QY 812	ccgtgcgtctgcgaagacacagc-gccgcagactgtgtgacatctgcgtctggagacattccg	870		
Db 311	CCGTGCGCTGCCAGACACAGCAGCCGAGACGTGTHGCACTTGCTTGGGAGCACTTCCGG	370		
QY 871	ccccgc-tgcaagtctcgt--ctgggggcccacaagacaagctgtcggaagttcaagccagc	927		
Db 371	ACCCGATGCAAGTTCGTATGGGGGCCACCAAGCAAGCTTCGGAGGTTCAAGCCAGC	430		
QY 928	tctctgcgcagcat-cgcgtcgaaggtgtccagaagaagtgtgagggagaaagcccgagga	986		
Db 431	TCTCTGCCCATACCGCTTCGAAGGCTCCAGGAAGTGTGAAGAGAAAGAACGCCGGGGA	490		
QY 987	ccccgacacagagcgacgaaccccaaggttcgaacctgttggccag-----agcatagcaa	1040		
Db 491	CCCCACACACAGGCGACACCCACACGCTGTGCACTGTGGGCCAGAGCATATCAACGGCT	550		
QY 1041	gggttggggcaggggtgtggcgaggggccccagacagagggcggtggagcccccgag---aagc	1097		
Db 551	GCCGGGACAGGTGGACAGACGGGACCCCAAGGCAAGAGCGTGGAGCCCCCAAGATGCCG	610		
QY 1098	gggaaccccttggagagagcgacaggggga---tgaggcagagggggccacgggggaagatagcc	1154		
Db 611	GGAGACATGCAAGAGAGCCACAGGGGACTGAGCGCAAGGCGCCCAAGCAACATCA	670		
QY 1155	ggag-----cccttaagcccccaagagagcgcaagaagagagcttga-----gtctgagcc	1204		
Db 671	GCAGGAGCCCATATTCGACAGCAAGAAAGAGCCACACACCAAAAGATGGCGCTGAGCC	730		
QY 1205	ggcgagagcagcccgccacagagccagggcccttcaagcttgcctcaaggttggagagatcg	1264		
Db 721	CGCGAGAGCATCG-CCACAGAGGCCCAAGGCTCAAGTGTCTCA-AGGTGCAAGAAAGCACG	788		
QY 1265	ctctgaattt 1274			
Db 789	CACGTGAAGT 798			

[illegible]

```

QY 346 cccaa-cggtgtttcatitgagacattcttcagaactgagagacaactatgacctct 404
Db 394 CCCAAGCGGCTGTCCATTGAGGACATTTTCAGAACCTGGACGACAACTATGACCTCCT 453
QY 405 tg-aggacaatcaactcctacatccagtggctgttccctcgcgagaaccagagtgaaact 463
Db 454 TGCAGGACATCACTCCTACATCCAGTGGCTGTCTCCTCTGCCAGACCCAGAGTGACT 513
QY 464 ggcattgccaagcccc-tcagctcaggga-gtccaggtgtttaaaagctccagagat 521
Db 514 GGCATGCCAAGCCCTTCACAGCTCAGGAGAGGTGAGAGGCTCCAGAGAT 573
QY 522 ccaggaacggctgttcgggacctagagctcat-gctgggtcttaaggtatcc-ggtg 579
Db 574 CCAAGAACGGCTGTTCGGGCTTACGAGCTCATAGCTGGGCTTCTACGGGATCCGGGCTG 633
QY 580 gaggacgagagcagc--ggcaggtggccgagcacagaactaccagaagcgctccaga 637
Db 634 GAGGACCGAGGACGGGGCCAGGTGGCCCGACACAGAACTACAGAAAGCGCTTCCAGA 693
QY 638 acctgaactggcgagccaca 658
Db 694 GCCTAACTGGGACAGCACAAA 714

```

Search completed: October 12, 2000, 12:03:21
 Job time: 24189 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 18:22:31 ; Search time 563.62 Seconds
(without alignments)
1525.659 Million cell updates/sec

Title: US-09-431-843B-11

Perfect score: 2289

Sequence: 1 tagaattcagcgccgcgtga.....ccttcgtgagtggtggggcgcg 2289

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_36:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	559	24.4	789	20	X40044
2	550.4	24.0	802	20	X40045
3	303	13.2	375	20	V90541
4	84	3.7	114955	20	X53491
5	71.2	3.1	1925	20	X90924
6	71.2	3.1	114955	20	X53491
7	69	3.0	10732	21	A10594
8	68.6	3.0	789	20	X40044
9	67.2	2.9	799	19	V55831
10	67.2	2.9	5452	20	X90923
11	67.2	2.9	8705	20	Z23778
12	67.2	2.9	9600	19	V21683

Result No.	Score	Query Match	Length	DB ID	Description
13	67.2	2.9	10380	20	Z22248
14	67.2	2.9	10596	14	O51731
15	67.2	2.9	10596	17	T40348
16	67.2	2.9	10596	20	X15650
17	65.8	2.9	35100	20	V73802
18	65.8	2.9	137507	19	V19941
19	64.8	2.8	1542	10	N90025
20	62.6	2.7	390	13	O21833
21	62.6	2.7	390	14	O36859
22	62.4	2.7	795	19	V55830
23	61.8	2.7	8438	15	O73500
24	60.2	2.6	32207	20	V73805
25	60.2	2.6	137507	19	V19941
26	60	2.6	117213	19	V62176
27	58.8	2.6	1337	20	V17263
28	58.4	2.6	150	20	V64556
29	58.4	2.6	51259	18	X83007
30	58	2.5	1218	21	A02488
31	58	2.5	6225	20	X55273
32	57.4	2.5	6225	21	A34721
33	57.4	2.5	1593	21	A02504
34	56.2	2.5	2277	19	V13836
35	56.2	2.5	2277	19	V05372
36	56	2.4	16442	18	X83006
37	55.8	2.4	309	10	N90579
38	55.8	2.4	1091	16	O85844
39	55.6	2.4	768	18	T84841
40	54.8	2.4	768	20	X35871
41	54.8	2.4	2643	14	O39212
42	54.2	2.4	761	16	O89794
43	54	2.4	2109	20	Z07194
44	54	2.4	1075	15	O64676
45	52.6	2.3			

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description
1	559	24.4	789	20	X40044
2	550.4	24.0	802	20	X40045
3	303	13.2	375	20	V90541
4	84	3.7	114955	20	X53491
5	71.2	3.1	1925	20	X90924
6	71.2	3.1	114955	20	X53491
7	69	3.0	10732	21	A10594
8	68.6	3.0	789	20	X40044
9	67.2	2.9	799	19	V55831
10	67.2	2.9	5452	20	X90923
11	67.2	2.9	8705	20	Z23778
12	67.2	2.9	9600	19	V21683

DR WPI: 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67: Page 646; 787p; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 789 BP; 157 A; 259 C; 275 G; 93 T; 5 other; *
SQ

Query Match 24.4%; Score 559; DB 20; Length 789;
Best Local Similarity 86.7%; Pred. NO. 5.4e-93;
Matches 685; Conservative 0; Mismatches 40; Indels 65; Gaps 4:

QY 1413 ggttgatgaggtgtctggtggagacagtctgctggtgacagtgtgtgtgtccacacattggc 1472
Db 1 ggttgatgaggtgtctggtggagacagtctgctggtgacagtgtgtgtgtccacacattggc 60
QY 1473 ccttgccgggtccctccctccctccctccctccctccctccctccctccctccctccctcc 1532
Db 61 ccttgccgggtccctccctccctccctccctccctccctccctccctccctccctccctcc 120
QY 1533 tgagggagacacagaaagtctgaaaggggcccacaaaggttacccctggagagccatcgga 1592
Db 121 tgagggagacacagaaagtctgaaaggggcccacaaaggttacccctggagagccatcgga 180
QY 1593 gaacccagagccca----- 1606
Db 181 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 240
QY 1607 -----gacccagagacagctgaggggagcagcagccagagccatcgga 1652
Db 241 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 300
QY 1653 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 1712
Db 301 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 360
QY 1713 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 1772
Db 361 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 420
QY 1773 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 1832
Db 421 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 480
QY 1833 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 1892
Db 481 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 540
QY 1893 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 1952
Db 541 gaacccagagcccaagccagacagctgaggggagcagcagccagagccatcgga 600
QY 1953 agcagagttgcagagcagaggtgaggtctcttcccaagt--ctggggaagcctaaagga 2010
Db 601 agcagagttgcagagcagaggtgaggtctcttcccaagtcttgggaagcctaaagga 660

QY 2011 aagagag-tgcccgctggcgtctgtc 2069
Db 661 aagagagttgcccgctggcgtctgtc 720
QY 2070 gagctgcagcggcccccctccacagctctgcgtgacccgtgacccatgacccacagtg 2129
Db 721 gactntctgnggtctcccttaaggtttgtt--gtgacctgacacatgancacacagt 778
QY 2130 ctgacctct 2139
Db 779 ctgacctct 788

RESULT 2
X40045/c
ID X40045 standard; DNA: 802 BP.
XX
XX X40045;
AC
XX
DT 02-JUL-1999 (first entry)
DE
XX Prostate cancer associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX W09904265-A2.
PN
XX
PD 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
PF
XX
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDWIG-) INST CANCER RES.
PA
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obara Y, Old LJ;
PI Pfeurdschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI: 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67: Page 646; 787p; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 802 BP; 112 A; 248 C; 277 G; 158 T; 7 other;

QY	1662	ccccgccccgagacctctgcaaggagacg-aaccacacgagagcccatctgagga-----cc	1716
Db	703	ccccgccccgcaaaagacattctcaaaaggacaaagaccacacgaagaccatttcggaaagacc	644
QY	1717	ccagagccccagccccgcg-c-aggaacctacaaggga-tgagccagccgagagcccatctgaga	1774
Db	643	ccagagccccagccccgcgcaagagacctcaaaaggagattgacccagaccgagagcccatctgaga	584
QY	1775	ccccagagccccgcgccccgcgcaaggaactctgcaagggaagagccagccgagagcccatctgaga	1834
Db	583	ccccagagccccccccccccgcgacgtgcagaggggacacagccagccgagagcccatctgagana	524
QY	1835	ccccagagccccgcgccccgcgcaaggaactctgcaagggaagagccagccgagagcccatctgaga	1894
Db	523	ccccagagccccccccccccgcgacgtgcagaggggacacagccagccgagagcccatctgagana	464
QY	1895	ccccagagccccagccccgcgcaaggaactctacaaggatgagccagccaaagtcgagggaagcag	1954
Db	463	ccccagagccccagccccgcgacgtgcacaaaggatgagccagccaaagtcgagggaagcag	404
QY	1955	cagaatttgacagagcagagaggtgagatctctctgcagaagctctggagagccttaagaaag	2014
Db	403	cagaatttgacagagcagagaggtgagatctctctgcagaagctctggagagccttaagaaag	344
QY	2015	agtgacccgtgagagctctgctgtccctctgtccctgtctgcaaggagtcgaggagccctcgagact	2074
Db	343	agtgacccgtgagagctctgctgtccctctgtccctgtctgcaaggagtcgaggagccctcgagact	284
QY	2075	gtctgagagctccctcctaagctctgtctctgtctgtgacccgtgacccatgacccaagtcgtgc	2134
Db	283	gtctgagagctccctcctaagctctgtctctgtgtgacccgtgacccatgacccaagtcgtgc	224
QY	2135	ctctctgagggagccactatagcagccacacagaaagccgcgagagccctcagaggaagcccaag	2194
Db	223	ctctctgagggagccactatgac	164
QY	2195	ctctgacagaagcctcctctgagcctgtgtctgtcttccacacacacacacacacacacacacac	2254
Db	163	ctctgacagaagcctcctctgagcctgt	104
QY	2255	tctttgtaattgacccctctgagatgaggggagcgcg	2269
Db	103	tctttgtaattgaccccttntgagagtgagggggcgcg	69
RESULT 3			
ID	V90541	V90541 standard; cDNA: 375 BP.	
AC	V90541;		
XX	15-FEB-1999	(first entry)	
DE	EST clone BK517.		
XX	Human; secreted protein; expressed sequence tag; EST; haematopoiesis;		
XX	tissue growth; activin; inhibin; chemotaxis; chemokines; haemostatic;		
XX	receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour		
XX	gene therapy; ss.		
XX	Homo sapiens.		
XX	OS		
XX	PN		
XX	MO9845436-A2.		
XX	15-OCT-1998.		
XX	PD		
XX	10-APR-1998.	98WO-US06955.	
XX	DE		

Query Match	13.2%	Score 303	DB 20	Length 375
Best Local Similarity	100.0%	Pred. No. 9,7e-47		
Matches 303	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1987	gccaaagtctgggaagaccttaagaaagagtgcccgctcgctctgtctccctgtccc	2046	
Db	22	gccaaagtctgggaagaccttaagaaagagtgcccgctcgctctgtctccctgtccc	81	
QY	2047	tgcttcgaagggtctggggtcctccggagctgcgtgcgggtctccctcaagctctgtctcgtga	2106	
Db	82	tgcttcgaagggtctggggtcctccggagctgcgtgcgggtctccctcaagctctgtctcgtga	141	
QY	2107	cccgagaccatgatgaccacaagctgcgtgcgtcctctgtgggggacataagagccacagaa	2166	
Db	142	cccgagaccatgatgaccacaagctgcgtgcgtcctctgtgggggacataagagccacagaa	201	
QY	2167	gccgagaggtcctcagagggaagcccaaggctctgcagaaagcctctcgtgctgtgtctt	2226	
Db	202	gccgagaggtcctcagagggaagcccaaggctctgcagaaagcctctcgtgctgtgtctt	261	
QY	2227	cccccccaagctctccctctgcgcctcttcttgaattgaacctcttgagtgagggg	2286	
Db	262	cccccccaagctctccctctgcgcctcttcttgaattgaacctctctgagtgagggg	321	
QY	2287	cggt 2289		
Db	322	cggt 324		
RESULT	4			
ID	X53491/c			
XX	X53491 standard; DNA; 114955 BP.			
XX	X53491;			
XX	05-JUL-1999 (first entry)			
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.			
XX	Antisense oligonucleotide; multiple target; antisense treatment;			
XX	impaired respiration; inflammation; lung disease;			

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 OS Synthetic.
 PN MO913886-A1.
 PD 25-MAR-1999.
 PF 17-SEP-1998; 98WO-US19419.
 PR 09-JUN-1998; 98US-0093972.
 PR 17-SEP-1997; 97US-0059160.
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 XX Nyce JW;
 DR WPI: 1999-229400/19.
 XX
 PI New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 XX
 PS Disclosure: Page 37; 120pp; English.
 CC
 CC The specification describes antisense oligonucleotides (X52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of RNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences X55272-74. These multiple target
 CC oligonucleotides (specifically X55180-271) can be used for the antisense
 CC treatment of diseases and conditions. Typical diseases and conditions
 CC are those associated with impaired respiration and inflammation,
 CC including lung diseases, pulmonary vasoconstriction, inflammation,
 CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
 CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
 CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
 CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
 CC metastases, as well as all types of cancers which may metastasize or have
 CC metastasized to the lungs, including breast and prostate cancer.
 CC
 SO Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 3.7%; Score 84; DB 20; Length 114955;
 Best Local Similarity 36.6%; Pred. No. 6.2e-07;
 Matches 421; Conservative 50; Mismatches 666; Indels 13; Gaps 4;

QY 757 gtgcgagcgagctgcgagcgagctgcgagagtgcccttgactacttcattgcgcgtg 816
 Db 106404 GAGCCAGGGGCCCNHNNHNSGCGGCGCGGCGAGCCAGGCGCCCNHNNHNSCGGCG 106345
 QY 817 cgtcgcgacacacagcgccgagctggtgcacttcgcttgaggagactccgagccgcg 876
 Db 106344 CGCGGGCGCGAGCCAGGGGCCCNHNNHNSGCGGCGCGGCGCGAGCCAGGCGCCCN 106285
 QY 877 tcgaagttcgtctggggcgcccaagaagctgcgaggttcaagccagctctctgcgc 936
 Db 106284 HNNNAGCGCGGCGCGGCGCGAGCCAGGCGCCCNHNNHNSAAGCGCGGCGGCGCG 106225
 QY 937 catcgctcgaggttcgaagaagtggaggaagaagaccccgaggagcccgacac 996

Db 106224 GAGCCAGGGGCCCNHNNHNSAAGCGGCGCGGCGCGAGCCAGGCGCCCNHNNHNSG 106165
 QY 997 gaggccagaccacaggtctgacctgtggtccagataagcaaggttggggcaggtg 1056
 Db 106164 CAAGCGGCGCGCGGCGCGAGCCAGGCGCCCNHNNHNSGCAAGCGGCGCGGCGCG 106105
 QY 1057 gacgaagggcccgccagccagagcgttgagcccaagatgtggagaccccttgagaagagc 1116
 Db 106104 GAGCCAGGGGCCCNHNNHNSGGCGAGCGGCGCGGCGCGAGCCAGGCGCCCNHNN 106045
 QY 1117 caggggagtgaagcaggggagccagggagataagtcgagcccttaagcccaagag 1176
 Db 106044 NSCGGCGCAAGCGCGGCGCGGCGCGAGCCAGGCGCCCNHNNHNSGGCGCGCAAGCGGCG 105985
 QY 1177 agc-aagaagaggaagcttgaagctgaagcggcgaggaagcagccgcaagaagagcc 1235
 Db 105984 CGCGGCGCGCGAGCCAGGCGCCCNHNNHNSGGCGGCAAGCGGCGCGCGCGCGAGCCA 105925
 QY 1236 tcagagttccctcagaggttggagaaagatcgtctgaatttggaggggtgtgtccctcagcca 1235
 Db 105924 GGGGCGCCCNHNNHNSGGCGGCGAGCCGCGGCGCGGCGCGAGCCAGGCGCCCNHNN 105865
 QY 1296 gggcag-----cctcagagcgggagcccaagagatggcggtcagagacccttgaggagcag 1351
 Db 105864 CGGCGCGCGCAAGCGCGGCGCGGCGCGAGCCAGGCGCCCNHNNHNSCGGCGCGCGCAAG 105805
 QY 1352 tcgagccttcgcccacaccccttgaggagcaggttggccgacaaggtga----ggaagcgg 1407
 Db 105804 CGGCGCGCGCGCGCGAGCCAGGCGCCCNHNNHNSGGCGGCGCGAGCCAGGCGCGCGG 105745
 QY 1408 aggaagtgatgaaggtgtgcctggggagacgtctgcgtgagcagttgtgtgtccagacc 1467
 Db 105744 GCGCGAGCCAGGCGCGCCCNHNNHNSGGCGGCGCGGCGCAAGCGGCGCGCGGCGCGAGCCA 105685
 QY 1468 ttggccttgcgggtgtcccttgcctccatcggggagaccacaaagcttgagacacagtgaagac 1527
 Db 105684 GGGGCGCCCNHNNHNSGGCGGCGCGGCGCGCAAGCGGCGCGCGGCGCGAGCCAGGCGCCCN 105625
 QY 1528 ggggttgggggagacaggaaggttcgaacggtggcccaagaaggtacccctggagagccca 1587
 Db 105624 HNNNSCGGCGCGGCGGCGAGCGGCGCGGCGCGAGCCAGGCGCCCNHNNHNSCGG 105565
 QY 1588 tcgagaccacacagcgcacagccagcagagacctcaggggagcggagccagccagagccca 1647
 Db 105564 GCGCGCGCGGCGAGCGGCGCGGCGCGAGCCAGGCGCGCCCNHNNHNSVCGGAGCGCV 105505
 QY 1648 tcgagaccacacagcgcacagcccgccgagcagaccttcagggagagcagccagccagagccca 1707
 Db 105504 GCG-----CCCNHNNHNSCGGCGCGGCGCCCNHNNHNSCGGCGCGCGCGG 105449
 QY 1708 tcgagaccacacagcgcacagcccgagcagagaccttaagggatggagccagccagagagccca 1767
 Db 105448 CGGAGHNNHNSCGGCGCGGCGCGCCCNHNNHNSCGGCGCGGCGCGGAGGAGGAGGCGG 105389
 QY 1768 tcgagaccacacagcgcacagcccgccagagaccttcagggagcggagccagccagagccca 1827
 Db 105388 CGGCGGAGHNNHNSCGGCGCGGCGCGGCGCGCCCNHNNHNSCGGCGCGGCGCGGCGGCGG 105329
 QY 1828 tcgagaccacacagcgcacagcccgccagagaccttcagggagagcagcagccagagagccca 1887
 Db 105328 SCGGCGCGCGGCGCGGCGCCCNHNNHNSCGGCGCGGCGCGGCGGCGGAGGAGGAGGCGG 105269
 QY 1888 tcgagagacc 1897
 Db 105268 CGGCGGCGCG 105259

RESULT 5
 ID X90924
 XX X90924 standard; DNA: 1925 BP.
 AC X90924;
 XX

DT	17-JAN-2000	(first entry)
XX		
DE	Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.	
XX		
KW	Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid PCMEBNA; EBNA 1;	
KW	episome; transfection; origin of replication; EBV oriP; receptor;	
KW	eucaryotic host cell; recombinant cell line; ion channel; gene therapy;	
KW	multiple gene expression; transporter protein; transcription factor;	
KW	adhesion molecule; antisense therapy; gene amplification;	
KW	cell immortalisation; ds.	
XX		
OS	Epstein-barr virus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1925
FT		/*tag= a
FT		/product= "EBNA 1"
FT		/transl_except= (pos:799..800, aa:Gly)
FT		/note= "The sequence is described throughout the
FT		specification as being 1926 nucleotides long, but a
FT		sequence of only 1925 bp has been given in figure 2"
XX		
PN	M09947647-A1.	
XX		
PD	23-SEP-1999.	
XX		
PX	12-FEB-1999; 99WO-US03307.	
PF		
PR	18-MAR-1998; 98US--0040961.	
PR	06-AUG-1998; 98US-0130114.	
XX		
PA	(PHAR-) PHARMACOEPIA INC.	
XX		
PI	DamaJ BB, Horlick RA, Robbins AK;	
XX		
DR	WPI; 1999-610610/52.	
DR	P-PSDB; YZ8843.	
XX		
PT	New method for expressing genes from recombinant eukaryotic cells,	
PT	useful for gene therapy -	
XX		
PS	Claim 24; Fig 2; 86pp; English.	
XX		
CC	The present sequence is a DNA encoding Epstein Barr Virus Nuclear	
CC	Antigen 1 (EBNA 1), which is obtained from commercially available	
CC	plasmid PCMEBNA. EBNA 1 protein is used to stably maintain episomes	
CC	containing EBV origin of replication (oriP) and a gene encoding	
CC	protein or RNA of interest. Eucaryotic host cells expressing EBNA 1	
CC	protein are transfected with these episomes to produce recombinant	
CC	cell lines expressing multiple genes of interest. This provides a	
CC	rapid and reliable method of stably expressing multiple genes in	
CC	transfected cells. The episomes are useful in the transfection of genes	
CC	encoding receptors, transporter proteins, ion channels, adhesion	
CC	molecules and transcription factors. The episomes carrying desired genes	
CC	can also be used to transfect cells in gene therapy, antisense therapy,	
CC	for gene amplification, cell immortalisation, etc.	
XX		
SQ	Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;	

Query Match	3.1%;	Score 71.2;	DB 20;	Length 1925;
Best Local Similarity	48.1%;	Pred. No. 9e-05;		
Matches 233;	Conservative 0;	Mismatches 248;	Indels 3;	Gaps 1;

OY	956 ggaagctcggaggagaagaagcccccgggagaccgcacagagccacgcccgatc	1015
Db	324 ggcacagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	383
OY	1016 ggaacctggggcagcagactaacagagtggtg--ggcacaggttggacgagggccccagc	1072
Db	384 agggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	443
OY	1073 cacggagccttgagagccccaagatgcgcggagccccccttgagagagcagcagcaggggatgagcag	1132

[illegible]

XX Cancer associated antigen; diagnosis: research; treatment: human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN WO9904265-A2.
XX
XX
XX 28-JAN-1999.
PD
XX
XX 15-JUL-1998; 98WO-US14679.
PF
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA
XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI Pfriendschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI: 1999-132448/11.
DR
XX
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT Isolated using sera from cancer patients, used to develop products
PI for the diagnosis, monitoring or treatment of cancers
XX
XX
XX Claim 67; Page 646; 787pp; English.
PS
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
CC
XX
XX Sequence 789 BP; 157 A; 259 C; 275 G; 93 T; 5 other;
SQ

DB 294 CTGGCTCTCCCTGCAGAGTCTCTGCTGGCGGCGGCTGGGGTCTCCAGATGGCTCTCGG 225
QY 1898 cagcccccccgccgagacacacaggaatgagccag 1936
DB 224 CTGGCTCTCCCTGCAGAGTCTCTGCTGGCGGCGGCGGCTGGGGTCTCCAGATGGCTCTCGG 186
RESULT 9
ID V55831 standard; DNA; 799 BP.
V55831
XX
XX V55831.
AC
XX
XX 18-NOV-1998 (first entry)
DT
XX
XX Nucleotide sequence of the stabilising sequence-encoding insert.
DE
XX
XX Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW Ikappab regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrg therapy; protease;
KW cancer; pathological condition; ss.
XX
XX Epstein-Barr virus.
OS
XX
XX W09822577-A1.
FN
XX
XX 28-MAY-1998.
XX
XX
XX 17-NOV-1997; 97WO-IB01508.
PF
XX
XX 25-JUN-1997; 97US-0048945.
PR
XX 15-NOV-1996; 96US-0030986.
XX
XX (MASU/) MASUCCI M G.
PA
XX
XX Masucci MG;
PI
XX
XX WPI: 1998-312463/27.
DR
XX
XX
XX New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PI a peptide sequence containing glycine repeats
XX
XX
XX Disclosure; Fig 4B; 120pp; English.
PS
XX
XX This is a nucleotide sequence of the stabilising sequence-encoding
CC insert. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula (Gly)_x(Gly)_y(Gly)_z where Gly_x, Gly_y, Gly_z are 1-6
CC sequential Gly residues and x, y, z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. x, y and z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an Ikappab regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
XX
XX Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;
SQ

Query Match 2.98; Score 67.2; DB 19; Length 799;
Best Local Similarity 47.7%; Pred. No. 0.00044;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;


```

XX OS Synthetic.
XX PN W09950457-A1.
XX PD 07-OCT-1999.
XX PF 28-MAR-1999; 99MO-US06742.
XX PR 28-MAR-1998; 98US-0079792.
XX PR 06-NOV-1998; 98US-0107504.
XX PA (UTAH ) UNIV UTAH RES FOUND.
XX PI Ruffner DE, Pierce ML, Chen Z;
XX DR WPI; 1999-610866/52.
XX PT Production of antisense libraries, used for identifying antisense
XX PT agents and for identifying target sites for antisense-mediated
XX PT inhibition of a selected gene -
XX PS Claim 16; Page 43-50; 63pp; English.
XX CC This invention describes a novel method for generating an antisense
XX CC library targeted to a selected RNA transcript. The methods can be used
XX CC for identifying antisense agents and for identifying target sites for
XX CC antisense-mediated inhibition of a selected gene. The use of a direct
XX CC library for target site selection significantly simplifies the screening
XX CC process, since only very small libraries need be prepared and assayed.
XX CC This sequence represents the vector pshuttle which is used in the method
XX CC of the invention.
SQ Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other:

```

```

Query Match 2.9%; Score 67.2; DB 20; Length 8705;
Best Local Similarity 47.7%; Pred. No. 0.00054;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

```

```

QY 956 ggaaggtggagggaggaagcccccggggagcccgacagagggccagaccagaggtc 1015
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7963 gccagagcagagcagagcagagggcgacagagggcgacagagggcgacagagcag 7904
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1016 ggaacctgtggccagagcagtaagcaaggttgg--gggcaagggtgagaggggcccacg 1072
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7903 AGGGGCGAGGAGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGAGGAGGCGGCG 7844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1073 cagcggggtgtgagggcccgatgtcggaacctgtgagagagagcagaggggtatgagcag 1132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7843 AGGAGCGAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGG 7784
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1133 ggggccaagggggaagataagcccgagcccttaagccccaagagagagcaagaggaagc 1192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7783 GCGAGAGCGAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGAGG 7724
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1193 tggagctgagccggcgagagcccgccacagagccagggcccttaagatgtgctcagaag 1252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7723 AGGGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGG 7664
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1253 tggagaagaatcgctctgaatttggaggggtgtgcccctcagagggcagagccctcagaagcg 1312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7663 AGGGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGG 7604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1313 ggaaccaggaagtgggcgggttcagagcccttggagggcagtgtagcccttgcggccaagccc 1372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7603 AGGGGCGAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGAGG 7544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1373 tgggggccaaggtgtgcacaaagtgtgaggaagcggaggaaggttgaataaggtgtgtggg 1432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7543 AGGGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGG 7484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
ID V21683 standard; DNA; 9600 BP.
AC V21683;
XX 17-AUG-1998 (first entry)
DE Vector plasmid pCMVkmITR-EPI.
XX Polynucleotide delivery; plasmid pCMVkmITR-EPI; vector;
XX gene therapy; vaccine; polycationic agent; ss.
XX Chimeric - Epstein-Barr virus.
OS Chimeric - Adeno-associated virus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
XX Key Location/Qualifiers
XX CDS 14..2594
XX misc_feature /product= "EBV nuclear antigen A"
XX /tag= a
XX /tag= b
XX /tag= c
XX /tag= d
XX /tag= e
XX /tag= f
XX /tag= g
XX /tag= h
XX /tag= i
XX /tag= j
XX /tag= k
XX /tag= l
XX /tag= m
XX /tag= n
XX /tag= o
XX /tag= p
XX /tag= q
XX /tag= r
XX /tag= s
XX /tag= t
XX /tag= u
XX /tag= v
XX /tag= w
XX /tag= x
XX /tag= y
XX /tag= z
XX /tag= aa
XX /tag= ab
XX /tag= ac
XX /tag= ad
XX /tag= ae
XX /tag= af
XX /tag= ag
XX /tag= ah
XX /tag= ai
XX /tag= aj
XX /tag= ak
XX /tag= al
XX /tag= am
XX /tag= an
XX /tag= ao
XX /tag= ap
XX /tag= aq
XX /tag= ar
XX /tag= as
XX /tag= at
XX /tag= au
XX /tag= av
XX /tag= aw
XX /tag= ax
XX /tag= ay
XX /tag= az
XX /tag= ba
XX /tag= bb
XX /tag= bc
XX /tag= bd
XX /tag= be
XX /tag= bf
XX /tag= bg
XX /tag= bh
XX /tag= bi
XX /tag= bj
XX /tag= bk
XX /tag= bl
XX /tag= bm
XX /tag= bn
XX /tag= bo
XX /tag= bp
XX /tag= bq
XX /tag= br
XX /tag= bs
XX /tag= bt
XX /tag= bu
XX /tag= bv
XX /tag= bw
XX /tag= bx
XX /tag= by
XX /tag= bz
XX /tag= ca
XX /tag= cb
XX /tag= cc
XX /tag= cd
XX /tag= ce
XX /tag= cf
XX /tag= cg
XX /tag= ch
XX /tag= ci
XX /tag= cj
XX /tag= ck
XX /tag= cl
XX /tag= cm
XX /tag= cn
XX /tag= co
XX /tag= cp
XX /tag= cq
XX /tag= cr
XX /tag= cs
XX /tag= ct
XX /tag= cu
XX /tag= cv
XX /tag= cw
XX /tag= cx
XX /tag= cy
XX /tag= cz
XX /tag= da
XX /tag= db
XX /tag= dc
XX /tag= dd
XX /tag= de
XX /tag= df
XX /tag= dg
XX /tag= dh
XX /tag= di
XX /tag= dj
XX /tag= dk
XX /tag= dl
XX /tag= dm
XX /tag= dn
XX /tag= do
XX /tag= dp
XX /tag= dq
XX /tag= dr
XX /tag= ds
XX /tag= dt
XX /tag= du
XX /tag= dv
XX /tag= dw
XX /tag= dx
XX /tag= dy
XX /tag= dz
XX /tag= ea
XX /tag= eb
XX /tag= ec
XX /tag= ed
XX /tag= ee
XX /tag= ef
XX /tag= eg
XX /tag= eh
XX /tag= ei
XX /tag= ej
XX /tag= ek
XX /tag= el
XX /tag= em
XX /tag= en
XX /tag= eo
XX /tag= ep
XX /tag= eq
XX /tag= er
XX /tag= es
XX /tag= et
XX /tag= eu
XX /tag= ev
XX /tag= ew
XX /tag= ex
XX /tag= ey
XX /tag= ez
XX /tag= fa
XX /tag= fb
XX /tag= fc
XX /tag= fd
XX /tag= fe
XX /tag= ff
XX /tag= fg
XX /tag= fh
XX /tag= fi
XX /tag= fj
XX /tag= fk
XX /tag= fl
XX /tag= fm
XX /tag= fn
XX /tag= fo
XX /tag= fp
XX /tag= fq
XX /tag= fr
XX /tag= fs
XX /tag= ft
XX /tag= fu
XX /tag= fv
XX /tag= fw
XX /tag= fx
XX /tag= fy
XX /tag= fz
XX /tag= ga
XX /tag= gb
XX /tag= gc
XX /tag= gd
XX /tag= ge
XX /tag= gf
XX /tag= gg
XX /tag= gh
XX /tag= gi
XX /tag= gj
XX /tag= gk
XX /tag= gl
XX /tag= gm
XX /tag= gn
XX /tag= go
XX /tag= gp
XX /tag= gq
XX /tag= gr
XX /tag= gs
XX /tag= gt
XX /tag= gu
XX /tag= gv
XX /tag= gw
XX /tag= gx
XX /tag= gy
XX /tag= gz
XX /tag= ha
XX /tag= hb
XX /tag= hc
XX /tag= hd
XX /tag= he
XX /tag= hf
XX /tag= hg
XX /tag= hh
XX /tag= hi
XX /tag= hj
XX /tag= hk
XX /tag= hl
XX /tag= hm
XX /tag= hn
XX /tag= ho
XX /tag= hp
XX /tag= hq
XX /tag= hr
XX /tag= hs
XX /tag= ht
XX /tag= hu
XX /tag= hv
XX /tag= hw
XX /tag= hx
XX /tag= hy
XX /tag= hz
XX /tag= ia
XX /tag= ib
XX /tag= ic
XX /tag= id
XX /tag= ie
XX /tag= if
XX /tag= ig
XX /tag= ih
XX /tag= ii
XX /tag= ij
XX /tag= ik
XX /tag= il
XX /tag= im
XX /tag= in
XX /tag= io
XX /tag= ip
XX /tag= iq
XX /tag= ir
XX /tag= is
XX /tag= it
XX /tag= iu
XX /tag= iv
XX /tag= iw
XX /tag= ix
XX /tag= iy
XX /tag= iz
XX /tag= ja
XX /tag= jb
XX /tag= jc
XX /tag= jd
XX /tag= je
XX /tag= jf
XX /tag= jg
XX /tag= jh
XX /tag= ji
XX /tag= jj
XX /tag= jk
XX /tag= jl
XX /tag= jm
XX /tag= jn
XX /tag= jo
XX /tag= jp
XX /tag= jq
XX /tag= jr
XX /tag= js
XX /tag= jt
XX /tag= ju
XX /tag= jv
XX /tag= jw
XX /tag= jx
XX /tag= jy
XX /tag= jz
XX /tag= ka
XX /tag= kb
XX /tag= kc
XX /tag= kd
XX /tag= ke
XX /tag= kf
XX /tag= kg
XX /tag= kh
XX /tag= ki
XX /tag= kj
XX /tag= kl
XX /tag= km
XX /tag= kn
XX /tag= ko
XX /tag= kp
XX /tag= kq
XX /tag= kr
XX /tag= ks
XX /tag= kt
XX /tag= ku
XX /tag= kv
XX /tag= kw
XX /tag= kx
XX /tag= ky
XX /tag= kz
XX /tag= la
XX /tag= lb
XX /tag= lc
XX /tag= ld
XX /tag= le
XX /tag= lf
XX /tag= lg
XX /tag= lh
XX /tag= li
XX /tag= lj
XX /tag= lk
XX /tag= lm
XX /tag= ln
XX /tag= lo
XX /tag= lp
XX /tag= lq
XX /tag= lr
XX /tag= ls
XX /tag= lt
XX /tag= lu
XX /tag= lv
XX /tag= lw
XX /tag= lx
XX /tag= ly
XX /tag= lz
XX /tag= ma
XX /tag= mb
XX /tag= mc
XX /tag= md
XX /tag= me
XX /tag= mf
XX /tag= mg
XX /tag= mh
XX /tag= mi
XX /tag= mj
XX /tag= mk
XX /tag= ml
XX /tag= mn
XX /tag= mo
XX /tag= mp
XX /tag= mq
XX /tag= mr
XX /tag= ms
XX /tag= mt
XX /tag= mu
XX /tag= mv
XX /tag= mw
XX /tag= mx
XX /tag= my
XX /tag= mz
XX /tag= na
XX /tag= nb
XX /tag= nc
XX /tag= nd
XX /tag= ne
XX /tag= nf
XX /tag= ng
XX /tag= nh
XX /tag= ni
XX /tag= nj
XX /tag= nk
XX /tag= nl
XX /tag= nm
XX /tag= no
XX /tag= np
XX /tag= nq
XX /tag= nr
XX /tag= ns
XX /tag= nt
XX /tag= nu
XX /tag= nv
XX /tag= nw
XX /tag= nx
XX /tag= ny
XX /tag= nz
XX /tag= oa
XX /tag= ob
XX /tag= oc
XX /tag= od
XX /tag= oe
XX /tag= of
XX /tag= og
XX /tag= oh
XX /tag= oi
XX /tag= oj
XX /tag= ok
XX /tag= ol
XX /tag= om
XX /tag= on
XX /tag= oo
XX /tag= op
XX /tag= oq
XX /tag= or
XX /tag= os
XX /tag= ot
XX /tag= ou
XX /tag= ov
XX /tag= ow
XX /tag= ox
XX /tag= oy
XX /tag= oz
XX /tag= pa
XX /tag= pb
XX /tag= pc
XX /tag= pd
XX /tag= pe
XX /tag= pf
XX /tag= pg
XX /tag= ph
XX /tag= pi
XX /tag= pj
XX /tag= pk
XX /tag= pl
XX /tag= pm
XX /tag= pn
XX /tag= po
XX /tag= pp
XX /tag= pq
XX /tag= pr
XX /tag= ps
XX /tag= pt
XX /tag= pu
XX /tag= pv
XX /tag= pw
XX /tag= px
XX /tag= py
XX /tag= pz
XX /tag= qa
XX /tag= qb
XX /tag= qc
XX /tag= qd
XX /tag= qe
XX /tag= qf
XX /tag= qg
XX /tag= qh
XX /tag= qi
XX /tag= qj
XX /tag= qk
XX /tag= ql
XX /tag= qm
XX /tag= qn
XX /tag= qo
XX /tag= qp
XX /tag= qq
XX /tag= qr
XX /tag= qs
XX /tag= qt
XX /tag= qu
XX /tag= qv
XX /tag= qw
XX /tag= qx
XX /tag= qy
XX /tag= qz
XX /tag= ra
XX /tag= rb
XX /tag= rc
XX /tag= rd
XX /tag= re
XX /tag= rf
XX /tag= rg
XX /tag= rh
XX /tag= ri
XX /tag= rj
XX /tag= rk
XX /tag= rl
XX /tag= rm
XX /tag= rn
XX /tag= ro
XX /tag= rp
XX /tag= rq
XX /tag= rr
XX /tag= rs
XX /tag= rt
XX /tag= ru
XX /tag= rv
XX /tag= rw
XX /tag= rx
XX /tag= ry
XX /tag= rz
XX /tag= sa
XX /tag= sb
XX /tag= sc
XX /tag= sd
XX /tag= se
XX /tag= sf
XX /tag= sg
XX /tag= sh
XX /tag= si
XX /tag= sj
XX /tag= sk
XX /tag= sl
XX /tag= sm
XX /tag= sn
XX /tag= so
XX /tag= sp
XX /tag= sq
XX /tag= sr
XX /tag= ss
XX /tag= st
XX /tag= su
XX /tag= sv
XX /tag= sw
XX /tag= sx
XX /tag= sy
XX /tag= sz
XX /tag= ta
XX /tag= tb
XX /tag= tc
XX /tag= td
XX /tag= te
XX /tag= tf
XX /tag= tg
XX /tag= th
XX /tag= ti
XX /tag= tj
XX /tag= tk
XX /tag= tl
XX /tag= tm
XX /tag= tn
XX /tag= to
XX /tag= tp
XX /tag= tq
XX /tag= tr
XX /tag= ts
XX /tag= tt
XX /tag= tu
XX /tag= tv
XX /tag= tw
XX /tag= tx
XX /tag= ty
XX /tag= tz
XX /tag= ua
XX /tag= ub
XX /tag= uc
XX /tag= ud
XX /tag= ue
XX /tag= uf
XX /tag= ug
XX /tag= uh
XX /tag= ui
XX /tag= uj
XX /tag= uk
XX /tag= ul
XX /tag= um
XX /tag= un
XX /tag= uo
XX /tag= up
XX /tag= uq
XX /tag= ur
XX /tag= us
XX /tag= ut
XX /tag= uu
XX /tag= uv
XX /tag= uw
XX /tag= ux
XX /tag= uy
XX /tag= uz
XX /tag= va
XX /tag= vb
XX /tag= vc
XX /tag= vd
XX /tag= ve
XX /tag= vf
XX /tag= vg
XX /tag= vh
XX /tag= vi
XX /tag= vj
XX /tag= vk
XX /tag= vl
XX /tag= vm
XX /tag= vn
XX /tag= vo
XX /tag= vp
XX /tag= vq
XX /tag= vr
XX /tag= vs
XX /tag= vt
XX /tag= vu
XX /tag= vv
XX /tag= vw
XX /tag= vx
XX /tag= vy
XX /tag= vz
XX /tag= wa
XX /tag= wb
XX /tag= wc
XX /tag= wd
XX /tag= we
XX /tag= wf
XX /tag= wg
XX /tag= wh
XX /tag= wi
XX /tag= wj
XX /tag= wk
XX /tag= wl
XX /tag= wm
XX /tag= wn
XX /tag= wo
XX /tag= wp
XX /tag= wq
XX /tag= wr
XX /tag= ws
XX /tag= wt
XX /tag= wu
XX /tag= wv
XX /tag= ww
XX /tag= wx
XX /tag= wy
XX /tag= wz
XX /tag= xa
XX /tag= xb
XX /tag= xc
XX /tag= xd
XX /tag= xe
XX /tag= xf
XX /tag= xg
XX /tag= xh
XX /tag= xi
XX /tag= xj
XX /tag= xk
XX /tag= xl
XX /tag= xm
XX /tag= xn
XX /tag= xo
XX /tag= xp
XX /tag= xq
XX /tag= xr
XX /tag= xs
XX /tag= xt
XX /tag= xu
XX /tag= xv
XX /tag= xw
XX /tag= xx
XX /tag= xy
XX /tag= xz
XX /tag= ya
XX /tag= yb
XX /tag= yc
XX /tag= yd
XX /tag= ye
XX /tag= yf
XX /tag= yg
XX /tag= yh
XX /tag= yi
XX /tag= yj
XX /tag= yk
XX /tag= yl
XX /tag= ym
XX /tag= yn
XX /tag= yo
XX /tag= yp
XX /tag= yq
XX /tag= yr
XX /tag= ys
XX /tag= yt
XX /tag= yu
XX /tag= yv
XX /tag= yw
XX /tag= yx
XX /tag= yy
XX /tag= yz
XX /tag= za
XX /tag= zb
XX /tag= zc
XX /tag= zd
XX /tag= ze
XX /tag= zf
XX /tag= zg
XX /tag= zh
XX /tag= zi
XX /tag= zj
XX /tag= zk
XX /tag= zl
XX /tag= zm
XX /tag= zn
XX /tag= zo
XX /tag= zp
XX /tag= zq
XX /tag= zr
XX /tag= zs
XX /tag= zt
XX /tag= zu
XX /tag= zv
XX /tag= zw
XX /tag= zx
XX /tag= zy
XX /tag= zz

```

Disclosure; Page 77-80; 100pp; English.

This polynucleotide comprises the DNA sequence of vector plasmid pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin of replication from plasmid pCEP4, a coding region for EBV nuclear antigen A from pCEP4, a pair of inverted terminal repeats from adeno-associated virus, a cytomegalovirus enhancer/promoter, a bovine growth hormone polyA sequence, and a kanamycin resistance selectable marker. Polynucleotides encoding polypeptides, such as erythropoietin or leptin, and ribozymes and antisense polynucleotides can be inserted into the vector. The vector is preferred for use in novel compositions and methods for improved polynucleotide delivery into cells. In these methods, polycationic

CC agents are used to increase the frequency of uptake of a
CC nucleic acid (see also V21684-86) into a cell. The polycationic
CC agent can condense with the nucleic acid and inhibit serum and/or
CC nuclease degradation of the nucleic acid. The nucleic acid can be
CC a vector, may express a therapeutic protein or a vaccinating viral
CC or cancer antigen, or is itself therapeutic (antisense or
CC ribozyme). The methods and compositions can be used in the gene
CC therapy of many diseases.

SD Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match	2.9%	Score 67.2;	DB 19;	Length 9600;
Best Local Similarity	47.7%	Pred. No. 0.00055;		
Matches 229; Conservative	0;	Mismatches 248;	Indels 3;	Gaps 1;

Qy	956	g g a a g g t c t g a g a g a a g a a g c c c c g g g a c c c g a c a a g a g g c c a a g a c c c a a g g t c	1015
Db	753	g g c a a g a a c a a g a g g a g g c a a g a g g c a a g a a g g c a a g a g g a a c a a g g g g	812

Qy 1016 ggaacctgtgagccagagcagaggtg--ggcaggtgtgacagagggcccaagc 1072

| | | | | | | | | | | | | | | |

Dd 813 aggggcagagcagagagagggcagggcagggcagggcagggcagggcagggc 872

Dy	1073	caagagcgtggaagcccaagatgcggaagccctgagagagcaaggggcatgagcag	1132
Dd	873	agagagcaagagagagggcagagaggggcaagagcaagagagagggcagagagaggg	932

Qy	1133	9999ccacggggaagatagggccggaagcccttaagccccaagaagagcagaagaaggaagc	1192
Db	933	ggcagaggaacagagagaggggacagaggaagagaggggacagagaggggacagagggg	992

Oy	1193	tggagctcagccgacgaggcagccccaagagcaggtgcctcagaag	1252
Dd	993	aggggacgaagggcgagggcgaggacgaaggaagggcagggacgaagggcgagg	1052

Qy	1253	tgagaagatcgtctgaattggaggggtgtgcctcagccagggcagcctcagacgg	1312
Db	1053	aggggcagagaggggcagagcagggggggcaggagcagagagaggggcagg	1112

Oy	1313	ggaacccagagaagtctggcgttcaggaccctctgggaagcagttcagcgcctctgcgcgcaaacccc	1372
Db	1113	aagvggcagagcagcagagagvggcagvggcagagagvggcagagcagcagagagvggcagagcagcag	1172

Qy	1373	tgggagccagggctggccagcaaaagctgaggaacgagagaaagctggatcctggg	1432
Db	1173	aggggcagagagggcagggcagggggggcagagagggcaggaaagcaggagggcagggag	1232

RESULT 13
222248

XX 222248; AC XX

XX Nucleotide sequence of pCEP4 vector.
DE
XX

KW calcium mobilization; nociceptin receptor; chemokine receptor; CCR3;
KW CCR2; interleukin 8 receptor type B; episomal expression vector; ss.
XX

XX WO9947921-A1.
PN
XX

XX	09-FEB-1999;	99WO-US02852.
PF		
XX		
10 MAR 1999		0000 00000001

PA (PHAR-) PHARMACOPEIA INC.

PI Damaj BB, Horlick RA, Robbins AK;

DR WPI; 1999-562218/47.

PT A method to enhance signal transduction responses involving cytoplasmic

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

XXIII
Pho 4

CC protein coupled receptor (G_iPCR) and with a gene coding for a G_i-a

CC assay a test compound for its effect on a G_iPCR. The method and cells

are useful for assaying agonist and antagonist compounds, which bind to the α_1 PCR which stimulate intracellular calcium mobilization, such as the

CC CCR2 and the interleukin 8 receptor type B. The present sequence

the presence of the vector
sequence in the
XX

Sequence 10380 BP; 2560 A; 2561 C; 2964 G; 2295 T; 0 other;

Query Match	2.9%	Score 67.2;	DB 20;	Length 10380;
Best Local Similarity	47.7%;	Pred. No. 0.00055;		
Matches 229; Conservative	0;	Mismatches 248;	Indels 3;	Gaps 1;

DY 956 ggaagctgagagaagaagaccgcggcgacccgacacagagcagcacccaggctc 1015
||| | | | | | | | ||| | | | | | | |
Db 6239 gccagagcagagaagagagccagagagccagagagccagagagccagagagcagag 6298

QY 1016 ggaactgtggcagagcatagcaaggttgg---ggcaggtgtgacgagggccaccagc 1072
| | | | | | | | | | | | | | | | | |
Db 6299 aggggcagagcagcaggaagggtgcaggaagggtgcaggaagggtgcaggaagggtgc 6358

Oy	1073	cacgagcgtggagccaccagatcgaggagcccttgagagagcacgagggagtgaagcag	1.132
Db	6359	agggacagagaggggcagagaaaggcagagcagagagaaggggcagagaaaggcagaagg	6418

QY 1133 ggggccagggagatagcgcgagcccttaagccccaagagagcaagaagagc 1192
 || | ||| | ||| ||| | ||| | ||| | |||
 Db 6419 ggcagagcagcagagaggggcagagcagcagagagggcgagaggggcagagcagag 6478

QY	1193	tgagagctgagccgagcgaggagcagccacagagccacagagctgctcagag	1252
Db	6479	agggcgaggaagggcaggaagggcaggaaggaaggaagggcaggaagcaggaagggcagg	6538

Qy	1253	tgagaaagatcgctctgaatttggaggggtgtgccctcagcagggcagctcaggaacgg	1312
Db	6539	aggggcaggaaggcaggaagcaggaaggcaggaagcaggaaggaaggcaggaaggcagg	6598

QY 1313 ggaaccaggaatgggcggtcaggaacctggggagcagtcagccctgcgcgaacccc 1372
| | | | | | | | | | | | | | | | | |
Db 6599 aggggcagggagccagggagggcaggaagcaggaaggggcaggaatcagg 6658

QY	1373	tgggagccagggtggcgcgaagaagtgaaagaagcggaggaagggtggaatgaaggtgcctgggg	1432
Db	6659	aggggcaaggaggggcaaggagcaggaggggcaaggaggggcaaggaggggcaaggagg	6718

RESULT 14
051731

XX
AC Q51731;
XX

XX
DE Plasmid pcISEBON for subcloning huhGF variants.
XX

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 12:03:21 ; Search time 7878.83 Seconds
(without alignments)
1796.262 Million cell updates/sec

Title: US-09-431-843B-11
Perfect score: 2289
Sequence: 1 tagaattcagcgccgctga.....cctctggagtgggggcg 2289

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_est1.*
2:	gb_est2.*
3:	gb_est3.*
4:	gb_est4.*
5:	gb_est5.*
6:	gb_est6.*
7:	gb_est7.*
8:	gb_est8.*
9:	gb_est9.*
10:	gb_est10.*
11:	gb_est11.*
12:	gb_est12.*
13:	gb_est13.*
14:	gb_est14.*
15:	gb_est15.*
16:	gb_est16.*
17:	gb_est17.*
18:	gb_est18.*
19:	gb_est19.*
20:	gb_est20.*
21:	gb_est21.*
22:	gb_est22.*
23:	gb_est23.*
24:	gb_est24.*
25:	gb_est25.*
26:	gb_est26.*
27:	gb_est27.*
28:	gb_est28.*
29:	gb_est29.*
30:	gb_est30.*
31:	gb_est31.*
32:	gb_est32.*
33:	gb_est33.*
34:	gb_est34.*
35:	gb_est35.*
36:	gb_est36.*
37:	gb_est37.*
38:	gb_est38.*
39:	gb_est39.*
40:	gb_est40.*
41:	em_estba.*
42:	em_estfun.*
43:	em_esthum1.*

44:	em_esthum2.*
45:	em_esthum3.*
46:	em_esthum4.*
47:	em_esthum5.*
48:	em_esthum6.*
49:	em_esthum7.*
50:	em_esthum8.*
51:	em_esthum9.*
52:	em_esthum10.*
53:	em_esthum11.*
54:	em_esthum12.*
55:	em_esthum13.*
56:	em_esthum14.*
57:	em_esthum15.*
58:	em_esthum16.*
59:	em_esthum17.*
60:	em_esthum18.*
61:	em_esthum19.*
62:	em_esthum20.*
63:	em_estin1.*
64:	em_estin2.*
65:	em_estin3.*
66:	em_estin4.*
67:	em_estom.*
68:	em_estov1.*
69:	em_estov2.*
70:	em_estpl1.*
71:	em_estpl2.*
72:	em_estpl3.*
73:	em_estpl4.*
74:	em_estpl5.*
75:	em_estro1.*
76:	em_estro2.*
77:	em_estro3.*
78:	em_estro4.*
79:	em_estro5.*
80:	em_estro6.*
81:	em_estro7.*
82:	em_estro8.*
83:	em_estro9.*
84:	em_estro10.*
85:	em_estro11.*
86:	em_estro12.*
87:	em_estro13.*
88:	gb_gss1.*
89:	gb_gss2.*
90:	gb_gss3.*
91:	gb_gss4.*
92:	em_gss1.*
93:	em_gss2.*
94:	em_gss3.*
95:	em_gss4.*
96:	gb_gss15.*
97:	gb_gss16.*
98:	gb_gss17.*
99:	gb_gss18.*
100:	gb_gss19.*
101:	em_gss5.*
102:	em_gss6.*
103:	em_gss7.*
104:	em_gss8.*
105:	em_gss9.*
106:	em_gss10.*
107:	em_gss11.*
108:	em_gss12.*
109:	em_gss13.*
110:	gb_gss5.*
111:	gb_gss6.*
112:	gb_gss7.*
113:	gb_gss8.*
114:	gb_gss9.*
115:	gb_gss10.*
116:	gb_gss11.*

This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Science Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.

No sl sequence available.
This clone (DKFZp762B143) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1..522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp762B143"
/clone_lib="762 (synonym: hmel2)"
/tissue_type="melanoma (Mewo cell line)"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI"
BASE COUNT 121 a 155 c 148 g 98 t
ORIGIN

Query Match 22.5%; Score 514; DB 14; Length 522;
Best Local Similarity 99.0%; Pred. No. 3.8e-94;
Matches 517; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 221 ccagaaactgcgagccagcagacatgttagttatcggtatcgacacatccggtatgg 280
Db 1 CCAGAAACTGGCGAGCCAGGAGGACATGTGTAGTATCGGCACAACTATCCGATCTGG 60
QY 281 tggacgagactgcaatgggacacgcgcaaacctcgagttctacagaaatgagatccgct 340
Db 61 TGGACGAGACTGCAATGGGACACGCGCAACACTGAGTTCTACAGAAATGAGATCCGCT 120
QY 341 tctgtcccaacggtgttctattgagacattcttcagaaacttcagacgacgaactatgacc 400
Db 121 TCCTGCCCAACGGCTGTTTCATTGAGGACATTTCTCAGAACTGGACGACAACTATGACC 180
QY 401 tcttgaggacaatcactcctacatcagtcagtgctgttctcctcgagaaaccaggaatga 460
Db 181 TCCTTGAGGACAACTACCTTACATCCAGTGGCTGTTCCTCGGAGAACCAAGAGTGA 240
QY 461 actggatgcgcaagccctcactcgctcagggaggtcgagtggtttttaaagctccccaggaga 520
Db 241 ACTGCGATGCCAAGCCCTCAGCTCAGGAGGTCGAGGTGTTTAAAGCTCCCGAGGAGA 300
QY 521 tccaggagcggctgtgtccgggacctacagctcatgctgtgggttctacggggtaccggtgg 580
Db 301 TCCAGGAGCGGCTGTTCGGGGCCTACGAGCTCATGCTGGGCTTCTACGGGATCCGGCTGG 360
QY 581 aggaccagggcacgggacgctggcgagcacagaaactacagaagcgtcttcagaaacc 640
Db 361 AGGACCGAGGACCGGGACAGGTGGCGGACGACAGAACTACCAAGCGCTTCAGAAACC 420
QY 641 tgaactggcgagcgcacacaaacctccgcacatcacacgcattcctaagtgcggtgtgagc 700
Db 421 TGAAGTGGCGAGCCACACAAACCTCCGCAATCACACGCATCCTCAAGTCGCTGGTGAGC 480
QY 701 tgagcctcgagcaactccagcgccactggtccgctttttcc 742
Db 481 TGGGCTCGAGCACTTCCAGGCGCGGTGGTCCGCTTCTTCC 522

RESULT 7
AW732857
LOCUS
DEFINITION
Similar to TR:096029 096029 7-60. ; mRNA sequence.
ACCESSION
VERSION
KEYWORDS
EST.

SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 526)
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
AUTHORS
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/lifresources.shtml
Seq primer: -40RP from Glibco
High quality sequence stop: 515.
FEATURES
Location/Qualifiers
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2963010"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 108 a 168 c 193 g 57 t
ORIGIN

Query Match 22.2%; Score 509; DB 24; Length 526;
Best Local Similarity 98.1%; Pred. No. 3.9e-93;
Matches 515; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1216 ccgcccacagacgagccctcagagtgctctcagagtggtgagaaagatcgctctgaattg 1275
Db 1 CCGCCACATTCGACGCCCTCAGAGTGCCTCAGAGTGGAGAGATCGCTCTGAATTG 60
QY 1276 gaggggtgtccctcagccgggagcagcctcagacgggggacccaggaagtggcggtcag 1335
Db 61 GAGGGGTGTCCTCAGCAGGCGAGCCTCAGGACGGGACCCAGGAAGTGGCGGTGAG 120
QY 1336 gacctgggagcagctgcagcctccgcaacccctggagccaggtggcgacgaag 1395
Db 121 GACCTGGGAGGAGTGTCAGCCTTCGCCCAACCCCTGGGAGCCAGGGTGGCGGACAG 180
QY 1396 gtgaggaagcggaggaaggtggatgaggtgctgggggacagtgctgctggcggtggcagtggt 1455
Db 181 GTGAGGAAGCGGAGGAGGTGGATGAGGGTGTCTGGGACAGTGTCTGGTGGCCAGTGGT 240
QY 1456 ggtgcccagacctggcccttgcgggtcccttcgcccctcaggggaccccaaggtcggga 1515
Db 241 GGTGCCCAGACCTTGGCCCTTGGCGGGTCCCTGCCCATCGGGGACCCCAAGGCTGGA 300
QY 1516 cacagtgaacgggttgagagagacacagaaggtcgagggggcccaagaaggtacc 1575
Db 301 CACAGTGAGAACGGGGTGTGAGGAGGACATAGGTGGAACGGGGCCCAAGAGGTACC 360
QY 1576 cctgggagcccatcgagaccccgagccagccagccagccagcagacctgcaggggacgagcca 1635
Db 361 CCTGGGAGCCCATCGGAGACCCCGAGGCCCGCCCGCCACGACGACCTGTCAGGGGACGAGCCA 420
QY 1636 gccgagagcccatcgagagaccccgagggcccgcccgccgaggaacctgcaggggacgagcca 1695

```
Db 421 GCCGAGAGCCCATCGGAGACCCCATGCCCGCAGCCGCGACGAGCTACAAGGGATGAGCCA 480
QY 1696 gccgagagcccatcgagagaccccgagggcccgagcccgagagacct 1740
Db 481 GCCGAGAGCCCATCGGAGACCCCGAGCCCGCCGCGCGGAGGAGCT 525

RESULT 8
LOCUS BE410591 591 bp mRNA EST 21-JUL-2000
DEFINITION 60130331f1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637634 5',
mRNA sequence.
ACCESSION BE410591
VERSION BE410591
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 591)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: ILCM337 row: f column: 03
High quality sequence start: 95
High quality sequence stop: 591.
FEATURES
source
1..591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3637634"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT 136 a 149 c 196 g 109 t 1 others
ORIGIN
Query Match 22.1%; Score 506.2; DB 35; Length 591;
Best Local Similarity 99.4%; Pred. No. 1.4e-92;
Matches 508; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 55 tccacctggagagacgacgagggatgcgaggaacgagagcagagctcgagagc 114
Db 81 TCCATGTGTGAGGAGGACGAGGAGGATCGGAGACGCGGAGGACGAGGACTCGAGGAC 140
QY 115 gccgagggcccgccgagggagcgagcgaggggacgagggacgagagagtcggaggag 174
Db 141 GCGAGAGCCCGCGCGGAGGAGCGGAGCGAGGGGACGAGGAGGAGGAGTCCGAGGAG 200
QY 175 ccgagggcgcgcgcccgagctcggttcagatccagatgacaggttcagaaactggcga 234
Db 201 CCGCGGCGGCGCGGCCACGCTCGTTCACATCCAGATGACAGGGTCCAGAAACTGGCGA 260
QY 235 gccacgagggagacatgtgtaggtatcggcacaaactatccggatctggtggaacgagactgc 294
```

```
Db 261 GCCACAGGAGACATGTGTAGGTATCGGCACAACTATCCGATCTGGTAACGAGACTGC 320
QY 295 aatggggacacgccaaacctgagttctacagaaatgagatccgcttcctgccaaacggc 354
Db 321 ATGGGGACACGCCAAACCTGAGTTCTACAGAAATGAGATCCGCTTCCTGCCCCAAGGC 380
QY 355 tgttcattgaggacattcttcagaactggacgacaactatgacctccttgaggacaat 414
Db 381 TGTTTCATTGAGGACATCTTCAGAACTGGACGACAACATATGACCTCCTTGAGGACAAT 440
QY 415 cactctaatccagtggtgtttctctctcgagagaccaggagtgactgcaatgcaag 474
Db 441 CACTCTATACCTCAGTGGCTGTTCTCTCGAGAACACGAGTGAACGTGCATGCCAAG 500
QY 475 ccctcacgctcagggaggtcgaggtgtttataaagctccagagatccagggcgctt 534
Db 501 CCCTCACGCTCAGGAGGTCGAGGTGTTTAAAGCTCCAGGAGATCCAGGAGCGGCTT 560
QY 535 gtcgggctcagagctcatgtctgggcttct 565
Db 561 GTCCGGCCTACGAGCTCATGCTGGGCTTCT 591

RESULT 9
LOCUS BE207411/c 551 bp mRNA EST 27-JUN-2000
DEFINITION ba66d10.x1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2905363 3',
similar to TR:096029 096029 7-60.;, mRNA sequence.
ACCESSION BE207411
VERSION BE207411.1 GI:8750809
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 551)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 434.
FEATURES
source
1..551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2905363"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT 96 a 175 c 180 g 100 t
ORIGIN
Query Match 20.9%; Score 478.4; DB 34; Length 551;
```


Tissue Procurement: DCTD/DTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing project Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bdrp/image/image.html Base Calling / Quality Scores: PHRED from University of Washington Genome Center. Vector Trimming: cross_match from University of Washington Genome Center PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley Drosophila Genome Project. University of Washington Genome Center: <http://www.genome.washington.edu> Polyadenylation: Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this CDNA insert was polyadenylated.

Plate: L1CM6 row: H column: 17
High quality sequence stop: 430.
Location/Qualifiers

FEATURES

source

1. .550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2821312"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8Kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 96 a 173 c 174 g 105 t 2 others
ORIGIN

Query Match 20.4%; Score 466.8; DB 21; Length 550;
Best Local Similarity 99.2%; Pred. No. 1.3e-84;
Matches 468; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1818 cgagagccatcgagagaccagcccccgcgaggaactcgagggagcagcagc 1877
Db 550 CGAGAGCCATCGAGAGACCAGCCAGGCCCCCGCCGCGCAGGACCTCGAGGGAGCAGCCAGC 491

Qy 1878 cgagagccatcgagagaccagcccccgcgaggaactcgagggatgagcagc 1937
Db 490 CGAGAGCCATCGAGAGACCAGCCAGGCCCCCGCCGCGCAGGACNTACAAGGGATGAGCCAGC 431

Qy 1938 caagcgggggagagcagagagtgagagcagagagtgagtgcttctgcaagtgctg 1997
Db 430 CAAGCGGGGAGGAGCAGAGAGTTGACGAGCAGAGAGTGAGTGCTTCTGCAAGTCTGG 371

Qy 1998 gaagccttaagaaagagtgcccgctggcgctttgttctctctgctcctgctgaggg 2057
Db 370 GAAGCCTTAAGAAAGAGTGCCGCTCGGCGTCTTGGTCCCTGCTCCCTGCTCAGGGG 311

Qy 2058 ctgggggctccgagagtgctgctggggctccctcaggtctctgctgacccgtgaccca 2117
Db 310 CTGGGGGCTCCGGAGCTGCTCGGGGCTCCCTCAGGCTCTGCTTCTGACCCGTGACCCCA 251

Qy 2118 tgacccacagtgctggctctctgtgggccaactatagcagccacagagccgagggcc 2177
Db 250 TGACCCACAGTGTCTGGCTCTCTGTGGGGCCACTATAGCAGCCACACAGAACCCGCGAGGCC 191

Qy 2178 ctgaggaagccccagggctgagaaagcctcctggtggtggtgtgttctcccccagcagc 2237
Db 190 CTCAGGGAAGCCCAAGGCGCTCAGAAAGCCTCTCTGGCTGGCTGTGTCTTCCCCACCCAGC 131

Qy 2238 tctccctgcccctgtcttctttaaattgaccttcttgagtggtggggggg 2289
Db 130 TCTCCCTGCGCCCTGTCTTGTAAATTGACCTTCTTGAGTGAGGGGGGGG 79

RESULT 12

BE398108

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE398108 529 bp mRNA EST 21-JUL-2000
601290328F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3620767 5',
mRNA sequence.
BE398108
BE398108.1 GI:9343473
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
1 (bases 1 to 529)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM293 row: g column: 08
High quality sequence start: 44
High quality sequence stop: 522.
Location/Qualifiers

FEATURES

source

1. .529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3620767"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8Kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 104 a 160 c 198 g 67 t
ORIGIN

Query Match 19.9%; Score 455.2; DB 35; Length 529;
Best Local Similarity 98.8%; Pred. No. 2.9e-82;
Matches 490; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

Qy 725 cactgtgccgtctctctctgagagagcgtgtgtcgcgcgaggagctcccgggggtgcggc 784
Db 35 CTCTGATCCGCTTCTTCTCTGAGGAGACACTGTGTGCGCGGAGCTCGCGGGGTGCGGC 94

Qy 785 agagtgccttgaggacttcttctgctgctgctgctgctgctgctgctgctgctgctgctg 844
Db 95 AGAGTGCCTTGAGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154

Qy 845 tgcacttcgctggagcacttcggcccgctgcaagttcgtctggtggggcccaagaca 904
Db 155 TGCACCTTCGCTGGAGACTTCCGGCCCCGCTGCAAGTTCGTCCTGGGGGGCCCCAAGACA 214

Qy 905 agctgcggaggttcaagcccgagctctctgccccctcgctcgagggctccaggaagtgg 964
Db 215 AGCTGCGGAGTTCAAGCCCGAGCTCTCTGCCCCATCGCTCGAGGGCTCCAGGAGTGG 274

Qy 965 agaggaaggaagccccgggggagccccagaccagagggccagccccaggttcgacctgtg 1024
Db 275 AGGAGGAAGGAAGCCCCGGGGGAGCCCCACACGAGGCGCAGACCCA-GGTCGGACCTGTG 333

```
QY 1025 gccagagcatagcaagggtggcagggtgagcaggggcccccagccagcgagcgtagg 1084
|||||
DB 334 GCCCAGAGCATAGCAAGGGTGGGGGCA-GGTGGACAGAGGGCCCCCAGCCAGCGAGCGTGG 392
|||||
QY 1085 agccccagatgcggggaacccctggagagagcagcagggggatgagcgagggggccacgggg 1144
|||||
DB 393 AGCCCCAGGATCGGGACCCCTGGAGAGGAGCCA-GGGGATGAGCAGGGGGCCACCGGG 451
|||||
QY 1145 aagatagccggagagccttaagcccaagaagagagcaagaagaagtgagctgaacc 1204
|||||
DB 452 AAGATAGCCGGAGGACCTTAAGCCCAAGAGAGCAAGAGAGGAAGCTGGAGCTGAGCC 511
|||||
QY 1205 ggcgggagcagcgcc 1220
|||||
DB 512 GCGGGAGCAGCGGCC 527

RESULT 13
AW246100 546 bp mRNA EST 07-JAN-2000
LOCUS 2821312.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821312 5',
DEFINITION mRNA sequence.
ACCESSION AW246100
VERSION AW246100.1 GI:6589093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 546)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Other ESTs: 2821312.3prime
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTP CDNA Library Preparation: Ling
Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
project Clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:
http://www.genome.washington.edu
Plate: L1CM6 row: H column: 17
High quality sequence stop: 470.
FEATURES
source
1. 546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2821312"
/clone_lib="NIH_MGC.7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
ORIGIN

Query Match 19.8%; Score 454.2; DB 21; Length 546;
Best Local Similarity 96.3%; Pred. No. 4.6e-82;

```
Matches 465; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1459 gccagaccttgcccttgccgggtgcgccttgccttcgccccatcgggggcaccacaaaggtggacac 1518
|||||
DB 9 GCCCAGACCTTGGCCCTTGGCGGTGCCCTTGCCTTCGCCCATCGGGGCACCCCAAGAGGTGGACAC 68
|||||
QY 1519 agtgagaaacgggttgagggagacacagaaggtcgaaacggggcccaagaaggtaccacct 1578
|||||
DB 69 AGTGAGAAACGGGGTTGAGGAGGACACAGAAAGTGCAGACGGGGCCCCAAAGAAAGGTACCCCT 128
|||||
QY 1579 gggagcccatcgagagaccccccagggcccccagcagcagcagcagcagcagcagcagcagcagc 1638
|||||
DB 129 GGGAGCCCATCGGAGACCCCGAGGCCCGCCAGCCAGCAGGACCTGCAGGGGACGAGCCAGCC 188
|||||
QY 1639 gagagcccatcgagagaccccccagggcccccagcagcagcagcagcagcagcagcagcagcagc 1698
|||||
DB 189 GAGAGCCCATCGGAGACCCCGAGGCCCGCCAGCCAGCAGGACCTGCAGGGGACGAGCCAGCC 248
|||||
QY 1699 gagagcccatcgagagaccccccagggcccccagcagcagcagcagcagcagcagcagcagcagc 1758
|||||
DB 249 GAGAGCCCATCGGAGACCCCGAGGCCCGCCAGCCAGCAGGACCTGCAGGGGACGAGCCAGCC 308
|||||
QY 1759 gagagcccatcgagagaccccccagggcccccagcagcagcagcagcagcagcagcagcagcagc 1818
|||||
DB 309 AAGACCCCATCGGAGACCCCGAGGCCCGCCAGCCAGCAGGACCTGCAGGGGACGAGCCAGCC 368
|||||
QY 1819 gagagcccatcgagagaccccccagggcccccagcagcagcagcagcagcagcagcagcagcagc 1878
|||||
DB 369 GAGAGCCCATCGGAGACCCCGAGGCCCGCCAGCCAGCAGGACCTGCAGGGGACGAGCCAGCC 428
|||||
QY 1879 gagagcccatcgagagaccccccagggcccccagcagcagcagcagcagcagcagcagcagcagc 1938
|||||
DB 429 GAGAGCCCATCGGAGACCCCGAGGCCCGCCAGCCAGCAGGACCTGCAGGGGACGAGCCAGCC 488
|||||
QY 1939 aag 1941
||
DB 489 GAG 491

RESULT 14
BE409569 1296 bp mRNA EST 21-JUL-2000
LOCUS 60129883F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629948 5',
DEFINITION mRNA sequence.
ACCESSION BE409569
VERSION BE409569.1 GI:9346019
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1296)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM317 row: e column: 21
High quality sequence start: 2
High quality sequence stop: 532.
FEATURES
Location/Qualifiers
1. 1296
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629948"
/clone_lib="NIH_MGC_21"
```

/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 327 a 416 c 389 g 164 t
ORIGIN

Query Match 19.4%; Score 443.6; DB 35; Length 1296;
Best Local Similarity 84.7%; Pred. No. 6.6e-80;
Matches 569; Conservative 0; Mismatches 89; Indels 32; Gaps 14;
Qy 515 agggatccaggagc-ggcttgcgggctacagctcatgctggtggtttctacgggac 573
Db 11 AGGAGATCCAGGAGTAGCTTGTCCGGGCTACGAGCTCATGCTGGGCTTCTACGGGATC 70
Qy 574 cgggtgagaccagagcagc-gcagcgtggccgagcacagaactaccagaagcgtt 632
Db 71 CGGTGGAGGACCGACCGACCGCGTGGCGGCGAGCACAGAACTACCAAGAGCGCTT 130
Qy 633 ccagaacctgaactggcgcagccacaacacctccgcacacacgcacatcctcaagtgc 692
Db 131 CCAGAACTGAACCTGGCGACGACCAACACCTCCGCATCACGCATCTCTCAAGTCGCT 190
Qy 693 gtgtgagctgagctcagcaactccaggccactggctcgttcttcttggaggagac 752
Db 191 GGGTGAAGTGGGCTCGAGCACTTCCAGGCGCCGCTGTGCGCTTCTCTGGAGGAGAC 250
Qy 753 gctggtcggcggagctgcgggggtgcgcagagtgccctgagctacttcatgttc-g 811
Db 251 GCTGTGCGCGGGAGTGTCCGGGGGTGCGGAGAGTGCCTGGAGTACTTCATGTTTCAG 310
Qy 812 ccgtgcctgcgcacacagc-gcgcgcagctgtgcacttcgctgagcacttcctcgg 870
Db 311 CCGTGGCTGCGGACACGAGGCGGACAGCTGTGTGCACTTCGGCTGGGAGCACTTCGG 370
Qy 871 cccgcg-tgcaagttcgt--ctggggggcccaagaagctgcggaggttcaagccagc 927
Db 371 ACCCGCATGCAAGTTCGTAATGGGGGCCCAACAGCAAGCTGCGGAGTTCAAGCCGAGC 430
Qy 928 tctctgcccacat-cgctcgggggtccaggaaggtgagaggaagaaagcccccgggga 986
Db 431 TCTGTGCCCCATACCGCTCGAGGGCTCCAGGAAGGTGGAGGAGGAGAGCCCGGGGA 490
Qy 987 cccgacacagccagccagccaggtcgcgacctgtgggcccag-----agcagacaa 1040
Db 491 CCGGACACAGGCGCCAGCCAGCTGCGGACCTGTGGGCGAGGAGCATAGCAACGGCT 550
Qy 1041 ggggtgggagcgggtggagcagggggcccgacagcagcgtgagagcccccagg---atgc 1097
Db 551 GCCGGCAGCAGTGGACAGACGGGACCCAGCCAGCGGAGCGTGGAGCCCCCAAGGATGCG 610
Qy 1098 ggggccccctggagagggagccaggggga---tgagggcagggggccacggggaagatagcc 1154
Db 611 GGAGCATATGAGAGGAGCCACGGGGAGTGGAGCGACCGGGGCCCGCCAGGGAACAATCA 670
Qy 1155 ggag-----cccttaagcccaagagagcagaagagagagcgtgga-----gctgagcc 1204
Db 671 GCAGAGCCCCATATACGCACCAACAGAGACGACGACGAGAAAGATGGCGCCCTGACCC 730
Qy 1205 ggcgggagcagccgcccacagcagcggcctcagagtgcctcagagtggtgagaagatcg 1264
Db 731 CGCGAGACCATCCG-CCACAGAGCCAGGCGCTCAGAGTGCCTCA-AGGTGCGAAGCAGC 788
Qy 1265 cctcgaattt 1274
Db 789 CACTGAACCTT 798

RESULT 15

BE391750

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE391750 757 bp mRNA EST 21-JUL-2000
601283311Fl NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3604994 5',
mRNA sequence.
BE391750
BE391750.1 GI:9337115
EST.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 757)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM252 row: f column: 03

High quality sequence start: 110

High quality sequence stop: 686.

FEATURES

source

1..757

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3604994"

/clone_lib="NIH_MGC_44"

/tissue_type="endometrium, adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 170 a 204 c 251 g 131 t 1 others

ORIGIN

Query Match 19.0%; Score 435.4; DB 35; Length 757;

Best Local Similarity 92.9%; Pred. No. 2.9e-78;

Matches 577; Conservative 0; Mismatches 31; Indels 13; Gaps 11;

Qy 51 cgactccacctgggagagcagcagcagcagcagcagcagcagcagcagcagcagc 110

Db 94 CAATTACACTGGGAGGAGGAGGAGGATGCGGATGACGCGGAGGAGGAGGAGT 153

Qy 111 ggaacggcgaagccgc 170

Db 154 GGACGGCAGGCG 213

Qy 171 ggaacggc 229

Db 214 GGAGCCCGGGCGGAGCGCGTCCCATCTCGTTCAGTCCAGATGACAGGGTCCAGAAACT 273

Qy 230 ggcagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 287

Db 274 GCGAGCCACGAGGAGCATGTGTACGATCGGCACAACTATCCGGATTCGTGTGGAACG 333

Qy 288 agactgcaatgggagacgacgacgacgacgacgacgacgacgacgacgacgacgacg 345

Db 334 AGACTGCAATGGGAGACACGCGCAACCTGAGTTTCTACGAGAAGGTGAGATCCGCTTCTG 393

```

QY 346 cccaa-cggctgtttcattgagacattcttcagaactgacggacaactatgacctct 404
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 394 CCCAAGCGCTGTTCATTGAGGACATTCTCAGAACTGGACGACAACTATGACTCCT 453
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 405 tg-aggacaatacactcctacatccagtggctgtttcctctcgagaaacaggagtgaact 463
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 454 TGCAGGACAATCACTCTACATCCAGTGGCTGTCTCTCTGCGAGACCCAGGAGTGAAC 513
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 464 ggcatgccaaagcccc-tcacgctcagga-ggtcaggtgttttaaaagctcccagagat 521
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 514 GGCATGCCAAGCCCCCTCAGCTCAGGGACGGTCGAGGCTTAGAGGCTCCAGGAGAT 573
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 522 ccaggagcggtgttcggggccctacagctcat-gctgggtctctacgggatcc-ggctg 579
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 574 CCAGGAGCGGCTTGTCGGGGCTACGAGCTCATAGCTGGCTTCTACGGGATCCGGGCTG 633
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 580 gaggaccggagcagc--ggcacggtggccgagacagaactaccagaagcgcttcacaga 637
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 634 GAGGACCGAGGACGCGGCCACGGTGGCCCGAGACAGAACTACCAGAAGCGCTTCCAGA 693
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 acctgaactggcgagccaca 658
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 694 GCCTAACTGGGCGAGCACCAAA 714

```

Search completed: October 12, 2000, 12:03:46
Job time: 24214 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 18:33:26 ; Search time 563.62 Seconds
(without alignments)
821.150 Million cell updates/sec

Title: US-09-431-843B-13

Perfect score: 1232

Sequence: 1 tagaattcagcgccgtga.....attcattgactttcgaaaa 1232

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_36.*
- 1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
 - 2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
 - 3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
 - 4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
 - 5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
 - 6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
 - 7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
 - 8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
 - 9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
 - 10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
 - 11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
 - 12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
 - 13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
 - 14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
 - 15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
 - 16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
 - 17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
 - 18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
 - 19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
 - 20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
 - 21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	112.6	9.1	802	X40045	Prostate cancer as
C 2	103.8	8.4	789	X40044	Prostate cancer as
C 3	94.6	7.7	375	V90541	EST clone BK517.
C 4	61.8	5.0	8438	V07500	DNA encoding Pseud
C 5	60	4.9	117213	V62176	HSV-2 strain SB5 C
C 6	58.4	4.7	51259	V83007	Partial mouse WRN
C 7	56.8	4.6	32207	V73805	KSHV LTR DNA (nucl
C 8	56.8	4.6	137507	V19941	KSHV long unique C
C 9	56	4.5	2277	V13836	Homo sapiens mamma
C 10	56	4.5	2277	V03372	Human telomerase p
C 11	56	4.5	16442	V83006	Partial mouse WRN
C 12	55.4	4.5	390	Q21833	Randomising oligon

13	55.4	4.5	390	14	Q36859	PCR primer for 5'
14	54.8	4.4	768	18	T84941	Human prostate pro
15	54.8	4.4	768	20	X35871	cDNA encoding a pr
16	54.2	4.4	2643	14	Q39212	CENP-B CDNA.. Homo
17	52.4	4.3	150	20	V64956	Mouse histone H2B
18	52.4	4.3	15672	12	Q10613	Rianodin receptor
C 19	52.2	4.2	49999	20	Z23900	Human LOBO homolog
C 20	52	4.2	766	20	Z17619	Human gene express
C 21	51	4.1	5321	20	V72945	Rat Munc13-1 encod
22	50.8	4.1	2109	20	Z07194	Human lung tumour
23	50.8	4.1	2301	19	V20445	Human c-trk oncoge
24	50.8	4.1	2301	21	A09303	Human Trk oncogene
C 25	50.6	4.1	53526	19	T94101	Human PKD1 gene.
C 26	50.6	4.1	53577	17	T18551	Human polycystic k
C 27	50.6	4.1	53577	19	T94108	Human PKD1 locus b
C 28	50.4	4.1	13330	20	Z23893	Human LOBO genomic
C 29	49.8	4.0	2277	19	V13834	Homo sapiens ambig
30	49.8	4.0	2277	19	V05370	Human telomerase p
31	49.4	4.0	795	19	V55830	FLGA insert stabil
32	49.4	4.0	799	19	V55831	Nucleotide sequenc
33	49.4	4.0	1925	20	X90924	Epstein Barr Virus
C 34	49.4	4.0	5452	20	X90923	Anti-sense strand
C 35	49.4	4.0	8705	20	Z23778	Vector pShuttle DN
36	49.4	4.0	9600	19	V21683	Vector plasmid pcm
37	49.4	4.0	10380	20	Z22248	Nucleotide sequenc
38	49.4	4.0	10596	14	Q51731	Plasmid pcISBON f
39	49.4	4.0	10596	17	T40348	Plasmid pcISBON f
40	49.4	4.0	10596	20	X15650	Nucleotide sequenc
41	49.2	4.0	1476	17	T08578	Zinc finger protei
C 42	49	4.0	4257	19	V68520	The nucleotide seq
C 43	49	4.0	4257	19	V10362	Infected cell prot
C 44	49	4.0	12001	16	Q76213	HSV L/ST region.
C 45	48.8	4.0	2784	18	T65102	Luciferase gene un

ALIGNMENTS

RESULT 1
X40045/c
ID X40045 standard; DNA; 802 BP.
XX
AC X40045;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14579.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
(LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX

```

DR WPI; 1999-132448/11.
XX
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; Page 646; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 802 BP; 112 A; 248 C; 277 G; 158 T; 7 other;
XX
XX
XX Query Match 9.1%; Score 112.6; DB 20; Length 802;
XX Best Local Similarity 91.5%; Pred. No. 5.4e-14;
XX Matches 151; Conservative 0; Mismatches 11; Indels 3; Gaps 3;
XX
XX QY 1064 gggaggcagaagcctgct-gcctggctgtgtctt-cccacccagctctcccctgcgcccc 1121
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 165 GGCGTCAGAGCCCTCTGGCCTGCGCTGCTGCTTCCCCACCAGCTWCCCTGCGCCCC 106
XX
XX QY 1122 tgtcttgttaatcgaccttcttgagcgggcgggcgggcggttgccctttcttag 1181
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 105 TGTCTTTGTAATTCACCTTNGAGTGGGGGCGGGCGAGGGC-TGCTTTTCTTAG 47
XX
XX QY 1182 tcgtatgcagaagggctttctgataaattcatttgacctt 1226
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 46 TCGTATCAGCAGCAGGCGCTTTCTGATAATTCATTTCACTTT 2
XX
RESULT 2
XX X40044
XX ID X40044 standard; DNA; 789 BP.
XX AC X40044;
XX XX
XX DT 02-JUL-1999 (first entry)
XX
XX XX Prostate cancer associated gene.
XX DE
XX DE
XX XX
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer; ss.
XX
XX OS Homo sapiens.
XX OS
XX W09904265-A2.
XX
XX XX
XX XX 28-JAN-1999.
XX
XX PF 15-JUL-1998; 98WO-US14679.
XX
XX XX
XX XX 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX
XX XX
XX PA (LUDWIG INST CANCER RES.

```

```

polyA_signal      8382..8387
                  /*tag= f
XX FT
XX FT
XX PN          US5352596-A.
XX XX
XX PD          04-OCT-1994.
XX XX
XX PF          11-SEP-1992;   92US-0945283.
XX PF
XX PR          11-SEP-1992;   92US-0945283.
XX PR
XX XX          (USDA ) US SEC OF AGRIC.
XX PA
XX PI          Cheung AK, Wesley RD;
XX XX
XX PI          WPI; 1994-316187/39.
XX DR
XX DR          P-PSDB; R60620.
XX XX
XX PT          New pseudorabies virus mutants for use in vaccine - having a
PT deletion and/or insertion in the early protein O gene or large
PT latency transcript gene
XX PT
XX PS          Disclosure; Column 15-30; 43pp; English.
XX CC
XX CC          Q73500 shows the Pseudorabies virus (PRV) large latency transcript
CC (LLT). The basic sequence is derived from PRV strain InfH and PRV
CC strain Ka. The LLT overlaps and is transcribed in the opposite
CC orientation with respect to the EP0 (early polypeptide 0) and the
CC immediately early gene (IE180). EP0 is nonessential for replicatio,
CC LLT is the only gene expressed during PRV latency, and the IE180
CC gene is absolutely necessary for PRV replication. However there are
CC 2 copies of IE180 in the genome. It is expected that PRV lacking one
CC of the IE180 copies is viable. Deletions in the non-overlapping
CC regions of these 3 genes will generate single deletion routants,
CC while deletions in overlapping regions will generate double deletion
CC mutants. The invention is concerned with the construction of attenuated
CC viruses which have a reduced ability to reactivate from latency. This
CC can be achieved by functionally disabling the expression of the EP0
CC gene, or by disrupting the synthesis of the LLT, or both.(See also
CC Q73501 and R60620-24)
XX CC
XX SQ          Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T; 0 other;

Query Match              5.0%; Score 61.8; DB 15; Length 8438;
Best Local Similarity    67.4%; Pred.No. 0.00071;
Matches      87; Conservative      0; Mismatches     42; Indels      0; Gaps

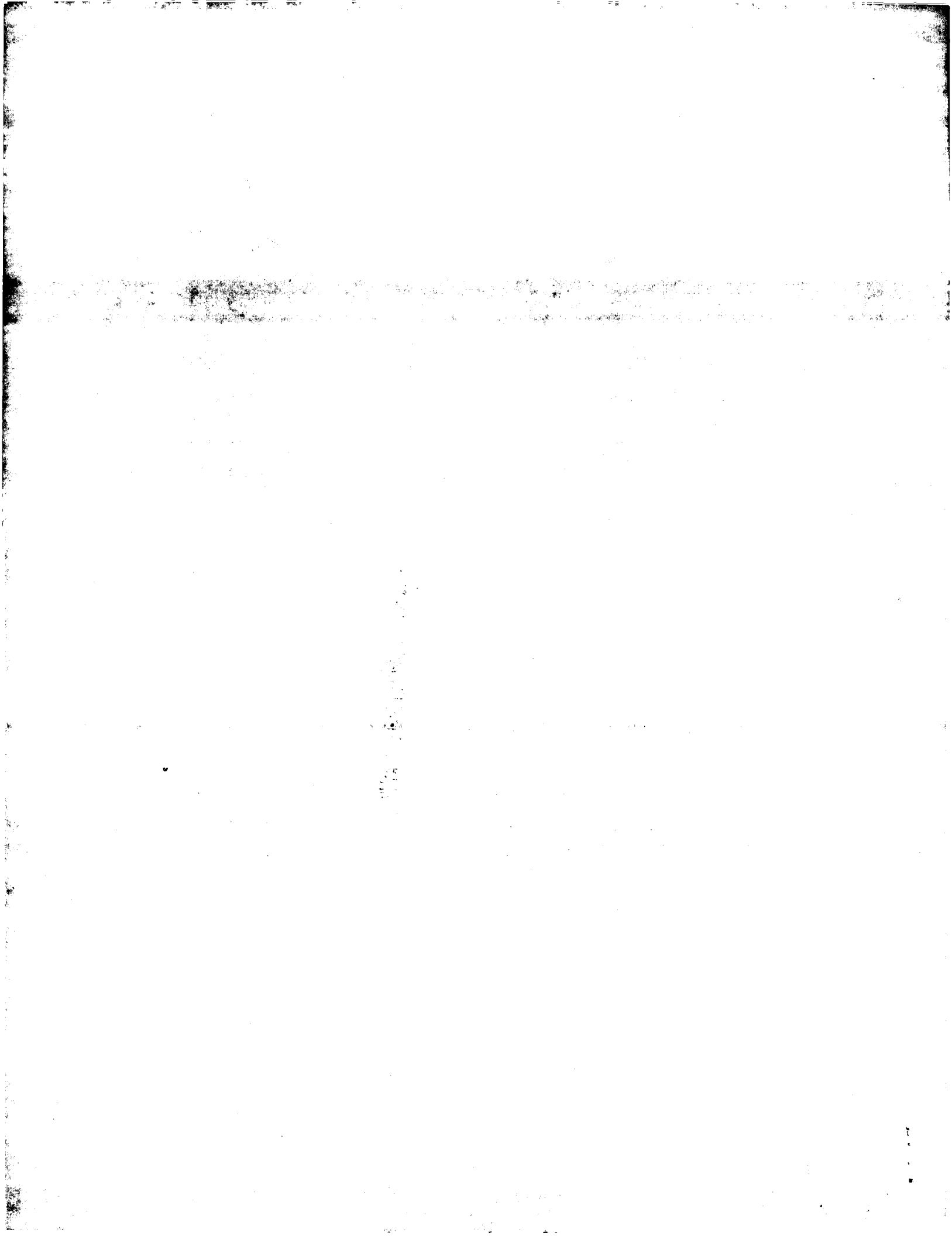
QY      64 ggagggaggacgaggaatcgaggacgcggaggacgaggaactcgagagcgcgcgagcc 123
Db       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5596 ggacgagcacgagagagacgagagagacgacgagagagagagagcgagcgccgcgcgcgc 123
QY      124 cgccgcgcgcgagggacgcgcgagcaggggacgagagacgaggaactcgagagcgccgcgcgc 183
Db       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5656 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 183
QY      184 ggcgcgcgcgc 192
Db       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5716 gtctgtgcc 5724

RESULT      5
V62176/C
ID V62176 standard; DNA; 117213 BP.
XX
XX V62176;
XX AC
XX DT
XX DT      13-JAN-1999 (first entry)
XX DE
XX DE      HSV-2 strain SB5 Contig ID 15 DNA sequence.
XX KW      HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor; ss

```

```
XX OS Herpes simplex virus type 2.
XX Location/Qualifiers
FH Key 755..1297
FT CDS /*tag= a
FT /product= "ORF#1 protein"
FT /note= "encoded protein shown in W72170"
FT 1170..2174
FT CDS /*tag= b
FT /product= "ORF#2 protein"
FT /note= "encoded protein shown in W72171"
FT 2229..2930
FT CDS /*tag= c
FT /product= "ORF#3 protein"
FT /note= "encoded protein shown in W72172"
FT complement (3130..3735)
FT /*tag= d
FT /product= "ORF#4 protein"
FT /note= "encoded protein shown in W72173"
FT complement (3802...6447)
FT CDS /*tag= e
FT /product= "ORF#5 protein"
FT /note= "encoded protein shown in W72174"
FT 6017..8482
FT CDS /*tag= f
FT /product= "ORF#6g protein"
FT /note= "encoded protein shown in W72250"
FT 6026..8482
FT CDS /*tag= g
FT /product= "ORF#6f protein"
FT /note= "encoded protein shown in W72249"
FT 6065..8482
FT CDS /*tag= h
FT /product= "ORF#6e protein"
FT /note= "encoded protein shown in W72248"
FT 6167..8482
FT CDS /*tag= i
FT /product= "ORF#6d protein"
FT /note= "encoded protein shown in W72247"
FT 6296..8482
FT CDS /*tag= j
FT /product= "ORF#6c protein"
FT /note= "encoded protein shown in W72246"
FT 6326..8482
FT CDS /*tag= k
FT /product= "ORF#6b protein"
FT /note= "encoded protein shown in W72245"
FT 6446..8482
FT CDS /*tag= l
FT /product= "ORF#6a protein"
FT /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT /transl_except= (pos: 7481..7486, aa: Ile)
FT 8457..9347
FT CDS /*tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in W72175"
FT complement (9604...11855)
FT /*tag= n
FT /transl_except= (pos: 11635..11636, aa: Ala)
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in W72176"
FT complement (11905...14508)
FT CDS /*tag= o
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in W72222"
FT complement (11905...14520)
FT /*tag= p
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in W72223"
FT 14399..15802
FT CDS /*tag= q
FT /product= "ORF#26 protein"
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in W72177"
FT complement (15996..16286)
FT /*tag= i
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in W72178"
FT complement (16202..18064)
FT /*tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in W72179"
FT complement (18105..19661)
FT /*tag= t
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in W72180"
FT complement (19415..20074)
FT /*tag= u
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in W72181"
FT 20155..21453
FT CDS /*tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in W72182"
FT complement (21326..22291)
FT /*tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in W72183"
FT complement (22546..24654)
FT /*tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in W72184"
FT 24684..25955
FT CDS /*tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in W72185"
FT complement (26295..27251)
FT /*tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in W72186"
FT complement (27630..31754)
FT /*tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in W72225"
FT complement (27630..31784)
FT /*tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in W72224"
FT complement (32067..32735)
FT /*tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in W72187"
FT 33140..34984
FT CDS /*tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in W72226"
FT 33386..34984
FT CDS /*tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in W72227"
FT complement (35205..37721)
FT /*tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in W72188"
FT complement (38058..39188)
FT /*tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in W72189"
FT 39090..39935
FT CDS /*tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in W72190"
FT 40216..41973
FT /*tag= ai
FT /product= "ORF#26 protein"
```

THIS PAGE BLANK (USPTO)



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 12:03:46 ; Search time 7878.83 Seconds
(without alignments)
966.796 Million cell updates/sec

Title: US-09-431-843b-13
Perfect score: 1232
Sequence: 1 tagaattcagcgccgctga.....attcattgacttcgaaaa 1232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss15:*
97: gb_gss16:*
98: gb_gss17:*
99: gb_gss18:*
100: gb_gss19:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: em_gss12:*
109: em_gss13:*
110: gb_gss5:*
111: gb_gss6:*
112: gb_gss7:*
113: gb_gss8:*
114: gb_gss9:*
115: gb_gss10:*
116: gb_gss11:*

QY 400 cctccttgagacatcactcctacatccagtggtgttctctctgcgagaccaggagt 459
|||||
Db 363 CTTCTTTCAGGACATCACTCCTACATCCAGTGGCTGTTCTCTCGCAGAACAGGAGT 422
QY 460 gaactggatgccaaagccctcagctcagggaggtcaggtgttttaaaactccagga 519
|||||
Db 423 GAATCGGATGCCAAGCCCTCAGCTCAGGAGGTGCGAGGTGTTTAAAGCTCCAGGA 482
QY 520 gat-ccaggagggc-ttgtcgggacctacagctcatgctgggcttctcaggatccgg 577
|||||
Db 483 GATCCAGAGCGGATGTCGGGCCCTACAGCTCATGCTGGGCTTCTACGGGATCCGG 542
QY 578 ctggaggaccagggcacgggca-cggtgggagcagacagaactaccagaagcgtt--- 633
|||||
Db 543 CTGGAGGACCGAGGACCGGCACCGGTGGCGGAGCAGACAGAACTACCAGAGCGCTTTCA 602
QY 634 cagaacctgaactggcgagcagcaacaacctccgcagcatcacagcatcctcaagtccgcg 693
| | |
Db 603 GAATCTGAATTTGGCGAGCCACAAACCTCCGCATCACACGCATTTCTCAAGTCGCTG 662
QY 694 tgtgagctgagcctcagca 713
| | |
Db 663 GTGACCTTGGGCTCGACGA 682

RESULT 2
LOCUS BE277227 868 bp mRNA EST 13-JUL-2000
DEFINITION 60117859F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3050950 5',
mRNA sequence.
ACCESSION BE277227
VERSION BE277227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 868)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/Drp
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM97 Row: h Column: 23
High quality sequence stop: 742.
Location/Qualifiers
1. 868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3050950"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
171 a 245 c 294 g 157 t 1 others

Query Match 46.9%; Score 577.6; DB 34; Length 868;
Best Local Similarity 87.9%; Pred. No. 9.8e-119;
Matches 752; Conservative 0; Mismatches 74; Indels 30; Gaps 10;
QY 52 cgaatcacactggagagagacgagggagatgcggagagacgcggagagagagactgc-g 110
|||||
Db 8 CGACTCCACCTGGGAGGAGGACGAGGAGGATGCGGAGGACGCGGAGGAGGAGGACTGCTG 67
QY 111 aggacggcgagggc-----gccggcgcgagggagacgcgggagcagggga 153
|||||
Db 68 AGGACGGCGATGGCTGCCCTGGCCCTGAGGTGACTGTGTGACTGCAGGGTGACTGAGGAC 127
QY 154 cgaggagagagagtcggagagacgcggcgggcgccagcagctcgttccagtcagaat 213
|||||
Db 128 GATGATGTCGATGATGATGCTGCTGGCGCTCGCCAGCTCGTTCAGTCCAGAAAT 187
QY 214 gacagggtccagaaac-tggcgagccacgagggacatgtaggtatcgcgcacaactatc 272
|||||
Db 188 GACAGGGTCCAGAAACTTGGCGAGCCACGAGGACATGTGTAGGTATCGGCACAACTATC 247
QY 273 cggatctggtggagacgagactgcaatgggagacgcgcaaacctcagttctcacagaatg 332
|||||
Db 248 CGGATCTGGTGGAAACGAGACTGCAATGGGGACACGCCAAACCTGAGTTTCTACAGAAATG 307
QY 333 agatccgccttctgcccacacggctgtttcattgaggagacattcttcagaactggcgagca 392
|||||
Db 308 AGATCCGCTTCTGCCCAACGGCTGTTTCATTGAGGACATTTCTCAGAACTGGACGACGA 367
QY 393 actatgacctcttgaggacaatcactcctacatccagtggtggttcttctctgcgagaac 452
|||||
Db 368 ACTATGACCTCTTGGAGGACAAATCACTCCTACATCCAGTGGCTGTTTCTCTCGGAGAAC 427
QY 453 caggagtaactggcagcatccaagccctcagctcagggaggtcaggtgtttaaagct 512
|||||
Db 428 CAGGAGTGAACTGGCATGCGCAAGCCCTCAGCTCAGGGAGGTGAGGTGTTTAAAGCT 487
QY 513 cccaggagatccagagagcgcttgcggggcctcagagctcagctggtggttctac-ggg 571
|||||
Db 488 CCCAGGAGATCCAGGAGCGGCTTGTCCGGGCTACGAGCTCATGCTGGGCTTCTACGGGG 547
QY 572 atccggtgagggaccagagcagcgggcaacggtggcgagcagacaactaccagaagcgc 631
Db 548 ATCCGGGTGGAGGACCGAGGACGCGGCGCGGTGGCGGAGCAGACAGAACTACCAAGAGCGC 607
QY 632 tt-cagaacctgaactggcgagcagccacaacacctccgcatcacacgcactcctcaagtcg 690
Db 608 TTCCAGAACTGAACTGGCGAGCTGCCGAGCCACAAACCTCCGCTATCACAGCATCTCTCAAGTGC 667
QY 691 ccgtgtgagctgagcctcgagcacttccagg--cgccactggttcgcttcttctctgga-g 747
Db 668 CTGGGTGAGCTGGGCTTCGAGCACTTCCAGGGCGCGCTGGGTCCGCTTCTTCTCTGGAGG 727
QY 748 gagacgtggtgcgg--cgggagctgcggggg--tgccgagagtgccctgg--actac 801
|||||
Db 728 GAGACGCTGGTTCGGGGCGGTGAGCTGCGGGGGGTGCGGGCAGAGGTGCTTGGGACTTAC 787
QY 802 ttcaigtccgcgtgcctgccacacacagcgcgcagctggtgcaactgcgcctggag 861
Db 788 TTATGTGTGCGCGCTGCGCTGCCGACACAGGCTGCCAGCTGTTGCCCTCGCCTTGGGA 847
QY 862 cacttcggcgcccgct 877
|||||
Db 848 GCCTTTCGGGCGCCCGT 863

RESULT 3
LOCUS BE387193
DEFINITION 601277089F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3618136 5',
mRNA sequence.
ACCESSION BE387193
VERSION BE387193
KEYWORDS EST.

/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 151 a 218 c 268 g 92 t
ORIGIN

Query Match 35.5%; Score 437.6; DB 34; Length 729;
Best Local Similarity 97.8%; Pred. No. 1.4e-87;
Matches 454; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 525 aggagcgcttgctcgccctacagctcgtggtgttcttaagggatccgggtggag 584
Db 1 AGGAGCGCTTGTCCGGGCTTACGAGCTCATGCTGGGCTTCTACGGGATCGGCTGGAGG 60
Qy 585 accgagcgacggcgacgttgccgagcacagaaactaccagaagcgtt-cagaacctga 643
Db 61 ACCGAGGCACGGGACCGGTGGCGGAGCACAGAACTACAGAGCGCTTCCAGAACCTGA 120
Qy 644 actggcgagcgcacaacacactccgcacacacgcacgtcctcaagtcgcgtgtgagctga 703
Db 121 ACTGGCGAGCCACACAAACCTCCGCATCACACGATCCTCAAGTCGCTGGGTGAGCTAG 180
Qy 704 gctctgagcaattccagcgccactggttcgcgttcttctctgagagacgctggtggcgc 763
Db 181 GCCTCGAGCACTTCCAGGCGCCACTGGTCCGCTTCTTCTGGAGGAGACGCTGGTGGCGC 240
Qy 764 gggagctccgggggtgcccagagtcgctggaactcctggaactcattatgctccctgcgtgcc 823
Db 241 GGGAGCTCCCGGGGGTGGCGGAGAGTGCCTTGGAGTCTTATGATGTCCTCCGCTGGCTGCC 300
Qy 824 gacacagcgccgcagctggtgcaacttcgctggtgagcaacttcggcccgctgcaagt 883
Db 301 GACACCAGCGCGCAGCTGCTGCACTTCGCTGGGACACTTCCGGGCCCGCTGCAAGT 360
Qy 884 tcgtctgggggccccaaagacagctgagggttcaagccagctctctgcgcgcatccgc 943
Db 361 TCGTCTGGGGGCCCCAAGACAAAGCTGCGGAGGTTCAAGCCAGCTCTCTGCGCATCGC 420
Qy 944 tcgagggctccagaaagtgcgagggagagacacctgcaggggac 987
Db 421 TCGAGGGCTCCAGGAAGTGTGAGGAGGAGGAAGACCCCGGGGAC 464

RESULT 7
BE391750
LOCUS BE391750 757 bp mRNA EST 21-JUL-2000
DEFINITION 601283311F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3604994 5', mRNA sequence.
ACCESSION BE391750
VERSION BE391750.1 GI:9337115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 757)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM252 row: f column: 03
High quality sequence start: 110
High quality sequence stop: 686.
Location/Qualifiers
source 1. .757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3604994"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 170 a 204 c 251 g 131 t
ORIGIN

Query Match 35.0%; Score 430.8; DB 35; Length 757;
Best Local Similarity 93.4%; Pred. No. 4.7e-86;
Matches 570; Conservative 0; Mismatches 27; Indels 13; Gaps 11;

Qy 52 cgaactccactggagggagacgagggatgoggagcgcgagggagggactgcga 111
Db 94 CAATTCACTGGGAGGAGGACGAGGAGATGCGGATGACGCGGAGGACGAGGACTGCGA 153
Qy 112 ggaacggagggcgccggcgaggggacgcgacgagggagggagacagagagtcgga 171
Db 154 GGAGCGGAGCGCCCGCGGAGGACGCGGAGGAGGACGAGGAGGAGTCCGA 213
Qy 172 ggagcccgggcgcg -cgcgcccgagctcgttccagtcacgaatgacaggggtccagaaact 230
Db 214 GGAGCGGCGGCGGACGCGTCCCATCTCGTTCAGTCCAGATCCAGAGGTCCAGAACT 273
Qy 231 ggagagcgaggggacatgtgta -ggatcgccacacactatccgga -tctgggtggagcg 288
Db 274 GGCAGGCGACGAGGAGACATGTGTACGGTATCGGCACAACTATCCGGATTCTGGTGGAGC 333
Qy 289 agactcaatggggacacgcgcaaacctgagtttctacagaaa--tgagatccgcttcctg 346
Db 334 AGACTGCAATGGGACACACGCCAAACCTGAGTTCTACAGAGGTGAGATCCGCTTCTCTG 393
Qy 347 cccaa -cggtgtttcattgagacattcttcagaactcttcagacgagggacgacaactatgacctct 405
Db 394 CCCAAGCGGCTGTTCATTTGAGGACATTTTCAGAACTGGAGCGGACAACTATGACCTCCT 453
Qy 406 tg-aggacaatcaactcctacacagtcagtggttctctcgagagaacccaggagtgaaact 464
Db 454 TGCAGGACAAATCACTCCTACATCCAGTGGGTGTCTCTCTCGGAGACCCAGAGTGAAC 513
Qy 465 ggaatgccaagcccc -tcacgctcagga -ggctcgaggtgttttaaagctcccgagagat 522
Db 514 GGATGCGAAGCCCTTCAGCTCAGGAGCGGTCCAGGGTGTAGAGGCTCCCGAGGAGAT 573
Qy 523 ccaggagcgcttgctcgccgctacagagctcat -gctgggctctctacggatcc -ggctg 580
Db 574 CCAGGAGCGGCTTGTCCGGGCTACGAGCTCATAGCTGGGCTTCTACGGGATCCGGGCTG 633
Qy 581 gaggacggaggaag --ggcaggtggcgagcacagaactaccagaagcgcttcagaa 638
Db 634 GAGGACCGAGGCGCGCCACCGGTGGCGGAGCACAGAACTACCAAGCGCTTCCAGA 693
Qy cctgaactgg 648
Db 694 GCCTAAGTGG 703


```
BASE COUNT      73 a   143 c   142 g   73 t   1 others
ORIGIN

Query Match
Best Local Similarity 33.4%; Score 411; DB 13; Length 432;
Matches 425; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 468 atgcaaacctcagctcagggagctgaggttttaaaagctccagagatccagg 527
Db 432 ATGCCAAGCCCTCAGCTCAGGAGGTCGAGGTGTTAAAGCTCNCAGGAGATCCAGG 373
QY 528 agcgtctgtccggcctacagctcagctggtgttctacgggacggtcgaggacc 587
Db 372 AGCGCTTGTCCGGCCACGAGCTCATGCTGGGCTTACGGGATCGGCTGGAGGACC 313
QY 588 gaggcacggcgctggtggccagacagaactacagaagcgtt-cagaacctgaact 646
Db 312 GAGGCACGGGACGGTGGGCGCTGCACAGACTACAGAAAGCGTTCCAGAACCTGAAC 253
QY 647 ggcgagccacacaacacctccgcacacacgcacatcctcaagtcgcgctgagctgagcc 706
Db 252 GCGCAGCCACAAACCTCCGCATCACACGCATCCTCAAGTCGCTGGTGAGCTGGGCC 193
QY 707 tcgagcacttcagggcgcactggtccgcttctctgagagagacgctggtgcgcggg 766
Db 192 TCGAGCACTTCCAGCGCGCTGTCGCTTCTTCTTGAGGAGACGCTGTCGCGCGG 133
QY 767 agctcccggggtgcgagagtcgctggactacttatcttccgctggtgcgcgac 826
Db 132 AGCTCCGGGGTGGCGAGAGTGCCTGGACTTCTATGTTCTGCCCTGCCCTCCGAC 73
QY 827 accagcgcgacgctggtgacttcgctgagagcacttcggcccgctgcaagtgcg 886
Db 72 ACCAGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13
QY 887 tctggggggccc 898
Db 12 TCTGGGGGGCCC 1

RESULT 10
AW250708      432 bp   mRNA      EST      07-JAN-2000
LOCUS      2822396.5prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2822396 5',
DEFINITION      mRNA sequence.
ACCESSION      AW250708
VERSION      AW250708.1 GI:6593701
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 432)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Other ESTs: 2822396.3prime
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/PTP CDNA Library Preparation: Ling
Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
project Clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:
http://www.genome.washington.edu
Plate: LUC99 row: E column: 21
```

```
FEATURES
Source
1..432
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2822396"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using 2AP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      80 a   141 c   143 g   68 t
ORIGIN

Query Match
Best Local Similarity 32.7%; Score 403; DB 21; Length 432;
Matches 417; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 506 aaagctccagagagatccagagcgttctccggcctacgagctcatctggcttc 565
Db 10 AAAAGCTCCACAGGAGATCCAGGAGCGGCTTGTCCGGGCTACGAGCTCATCTGGGCTTC 69
QY 566 tacgggatccggcgtggaggaccgagggcacggtggtggcgcagcagacagaactaccag 625
Db 70 TACGGGATCCGGCTGGAGGACCGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACG 129
QY 626 aagcgctt-cagaacctgaactggtgcgcagcacaacaacctccgcacatacagcgcctc 684
Db 130 AAGCGCTTCCAGAACCTGAACCTGGCGCAGCCACCAACCAACCTCCGCATCACACGATTC 189
QY 685 aagtcgcgctgtagctgagctcagcacttcagcgccactggtccgcttctctctg 744
Db 190 AAGTCGCTGGGTGAGCTAGGCTCGAGCACTTCCAGGGGCGCAGCTGGTCCGCTTCTCTG 249
QY 745 gagagacgctggtgcgcgggagctgcgggggtgcggcagagtgccctggactacttc 804
Db 250 GAGGAGAGCGCTGTCGCGGGGAGCTGCGGGGTGCGGAGAGTGCCTGGACTACTTC 309
QY 805 atgttcgcgtgctgcgcagacacacagcgcgcagctggtgcacttcgctggagcagc 864
Db 310 ATGTTCCCGCTGCGCTGCCGACACACGCGCGGCGAGCTGGTGCACCTTGGCTGGGAGCAC 369
QY 865 ttcggcccgctgcaagttctgctggtggggcccaagacagctgcgaggttcaagccc 924
Db 370 TTCGGCCCCCTCGAAGTTCGCTGTGGGGGGCCCCAAGACAGCTGGGAGGTTTCAAGCCC 429
QY 925 agc 927
Db 430 AGC 432

RESULT 11
BE391975      549 bp   mRNA      EST      21-JUL-2000
LOCUS      601285002F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3606840 5',
DEFINITION      mRNA sequence.
ACCESSION      BE391975
VERSION      BE391975.1 GI:9337340
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 549)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
```


Db 145 ACGGCCAAGCCTGTTCCAGTCTAGATGACAAGGTACCGAAACTGGCGTGCTATGCAGGA 204
Qy 247 catgtgtaggtatcgggaacaactatcgggatctcgggtggaacgagactgcaatggggacac 306
Db 205 CATGCAAGATACCGGCACAACTACCGGATTTGACAGATCAGACTGCAACGGAGACAT 264
Qy 307 gccaaacctgagttctctacagaatgagatcgcgtctcctgcccaacggtgtttcattga 366
Db 265 GTGCAACCTGAGCTTCTACAAAATGAGATCTGCTCCAGCCAAATGGCTTTCATTGA 324
Qy 367 ggacattctcagaactggcgggacaactatgacctccttgaggacaatcactcctacat 426
Db 325 GGACATTCTTCAGAACTGGAAGACAACATATGACCTCCTGGAGAGAATCACTCCTACAT 384
Qy 427 ccagtggctgtttcctctcgcagaaaccaggagtgaaactggcatgccaaagccctcacgct 486
Db 385 CCAGTGGCTGTTCCCTCGAGGGAACCAAGAGTGAACCTGGCATGGCATGCCAAGCCCCCTCACACT 444
Qy 487 caggaggtcaggtgttttaaagctccaggagatccaggagcgttgttccgggacctta 546
Db 445 GAAGGAGGTTGAGGCATTTAAAGCTCCAAGGAAGTCAGAGAGCGTCTTGTCGGGGCCTA 504
Qy 547 cgagctcatgtgggtctctacgggattccgggtcgtgagaccgaggaacgggcacggtggg 606
Db 505 TGAGCTCATGCTNGCTTCTATGGGATCCAACCTGAGGACCGGAACACAGGCGCGGTATG 564
Qy 607 ccgagcacagaactaccagaagcgct 632
Db 565 CCGTGCACAGAACCTCCAGCCACGCT 590

Search completed: October 12, 2000, 12:04:03
Job time: 24231 sec

THIS PAGE BLANK (USPTO)

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: moares@blue.veeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a Bonaldi poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized atrium at 16.5 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 forward.

FEATURES

source

Location/Qualifiers
1. 403
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-AAO-w1-g-01-0-UI"
/clone_1lb="UI-R-AAO"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AAO library is a non-normalized library constructed from 16.5 dpc rat atrium. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Jim Lin, Department of Biology, University of Iowa."

BASE COUNT 99 a 107 c 105 g 92 t
ORIGIN

Query Match 17.2%; Score 387; DB 11; Length 403;
Best Local Similarity 97.5%; Pred. No. 4.5e-75;
Matches 393; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1815 tgacacttggaacctagtgatgagtcagagagtgagcagagattgagctctgc 1874
DB 403 TGACACTTGGGAACCTAGTGATGACTCAGAGAGTGGCAAGATTGAGCCTTTGG 344
OY 1875 tgaaccccaagccttagagtgatcagctcactcagccacagtcaggggtttc 1934
DB 343 TGAACCCCAAGCCTTAGAGTGATTCAGTCTACTACGCCCAAGCGAGGGGCTTTT 284
OY 1935 tgaatcagagctctgcctgagctctcttcttggtgcccacagtgctctccctag 1994
DB 283 TGAGTCCAGAGCTTGGCTGAGCTTCTTCTGTCGCCCAAGTGGCTTCCTCCCTAG 224
OY 1995 tggatcagagtggtgagcagagagagagagagagagagagagagagagagag 2054
DB 223 TGGTACTGAGGTGGCCACAGAGGAGTGAAGCCCTCCCTCAGAGGAGCCAAAGGCT 164
OY 2055 tcaagacccctcctacacacatgtctcctcctcagctcagcctcagcctggttga 2114
DB 163 TCAGAACCTCTCTTACCTCAATGTGTCTCTCCCTCCAGCTCTTGAGGCCGCGTGTGA 104
OY 2115 tcaagacccctcagagtgagagagagagagagagagagagagagagagagag 2174
DB 103 TCAGACCTTAAGGGCTCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 44
OY 2175 ttctgataaactcttagactgtgcaaaaaaaaaaaaaa 2217
DB 43 TTCTGAATTAACCTTTAGACTTTGAAAAA 1

RESULT 13
AL120919

522 bp mRNA

EST

25-FEB-2000

DEFINITION DKFZP762B143.F1.762 (synonym: hmel2) Homo sapiens cDNA clone
DKFZP762B143.5', mRNA sequence.

ACCESSION AL120919
VERSION AL120919.1 GI:5926920
KEYWORDS EST.

SOURCE
ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 522)
AUTHORS Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE EST (Koehler, et al.)
JOURNAL Unpublished (1999)
CONTACT: Koehler K

MIPS
Am Kioferplatz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
genome project.
No 31 sequence available.
This clone (DKFZP762B143) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source

Location/Qualifiers
1. 522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP762B143"
/clone_1lb="762 (synonym: hmel2)"
/tissue_type="melanoma (Meko cell line)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 121 a 155 c 148 g 98 t
ORIGIN

Query Match 17.0%; Score 383.4; DB 14; Length 522;
Best Local Similarity 83.5%; Pred. No. 2.9e-74;
Matches 435; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

OY 309 ccgaactggtgctatgacagacatgcaagaatccggcacactaccggattgac 368
DB 2 CAGAACTGGCGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 61
OY 369 agatcaagactgaatggtgacatgtgcaacctggaacctctacaagaatgagctgctt 428
DB 62 GGAACGAGACTGCATAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
OY 429 ccagccaaatggggctctcagagacatcttcagaaactggaagaactatgacct 488
DB 122 CCTGCCCAAGGCGCTTTCATGTAGAGGACATTTCTCAAGAACTGAGACGACATGACCT 181
OY 489 cctggaagaagaatcaactcactacatccagtgcttctcctctgcygggaaccagagtgaa 548
DB 182 CTTGAGAGCAATACACTCTACATCCAGTGGCTTCTCTCCCTCCGAGAACCAAGAGTGA 241
OY 549 ctggcagcagcaaacccctcactcagagagagtgagagagagagagagagagagag 608
DB 242 CTGGCAATGCAGAGCCCTTCAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 301
OY 609 caagagagctctgtccgggagcctatgagcctatgagctgagctctatgaggtccacctga 668
DB 302 CCAAGAGCGGCTGTGCGGGGCTTACAGAGCTCAAGCTGAGGCTTCTACGGAGTCCGGCTGGA 361
OY 669 ggaacgggagcaggggtgtgtatgacgtgacagaaactccagcggcgtctcacatct 728
DB 362 GGACCGAGGAGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 421
OY 729 gaacagcagcagcagcaaacactgctattacacgcatctcaagtcattgagagct 788

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

AUTHORS Lindquester,G.J., Inoue,N., Allen,R.D., Castelli,J.W., Stamey,F.R.,
Dambaugh,T.R., O'Brian,J.J., Danovich,R.M., Frenkel,N. and
Pellett,P.E.
TITLE Restriction endonuclease mapping and molecular cloning of the human
herpesvirus 6 variant B strain 229 genome
JOURNAL Arch. Virol. 141 (2), 367-379 (1996)
MEDLINE 96195263
AUTHORS 4 (bases 64106 to 84963)
TITLE Lindquester,G.J., O'Brian,J.J., Anton,E.D., Greenmoyer,C.A.,
Pellett,P.E. and Dambaugh,T.R.
Genetic content of a 20.9 kb segment of human herpesvirus 6B strain
229 spanning the homologs of human herpesvirus 6A genes U40-57 and
containing the origin of DNA replication
JOURNAL Arch. Virol. 142 (1), 103-123 (1997)
MEDLINE 97300856
AUTHORS 5 (bases 106012 to 126166)
TITLE Lindquester,G.J., Greenmoyer,C.A., Anton,E.D., O'Brian,J.J.,
Pellett,P.E. and Dambaugh,T.R.
Comparison of a 20 kb region of human herpesvirus 6B with other
human beta herpesviruses reveals conserved replication genes and
adjacent divergent open reading frames
JOURNAL Arch. Virol. 142 (1), 193-204 (1997)
MEDLINE 97300864
AUTHORS 6 (bases 1 to 162114)
TITLE Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S., Inoue,N. and
Pellett,P.E.
Human herpesvirus 6B genome sequence: coding content and comparison
with human herpesvirus 6A
JOURNAL J. Virol. 73 (10), 8040-8052 (1999)
MEDLINE 99412318
AUTHORS 7 (bases 19696 to 22886)
TITLE Pellett,P.E.
Direct Submission
JOURNAL Submitted (26-JUL-1993) Herpesvirus Section, Centers for Disease
Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta,
GA 30333, USA
REFERENCE 8 (bases 64106 to 84963; 106012 to 126166)
AUTHORS Pellett,P.E.
Direct Submission
JOURNAL Submitted (02-OCT-1993) Herpesvirus Section, Centers for Disease
Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta,
GA 30333, USA
REFERENCE 9 (bases 1 to 162114)
AUTHORS Pellett,P.E., Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S.
and Inoue,N.
Direct Submission
JOURNAL Submitted (10-JUN-1999) Herpesvirus Section, Centers for Disease
Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta,
GA 30333, USA
REMARK Sequence updated by submitter
COMMENT On Aug 16, 1999 this sequence version replaced gi:405170 gi:405156
gi:305397.
FEATURES
source location/Qualifiers
1. 162114
/organism="Human herpesvirus 6B"
/strain="229"
/db_xref="taxon:32604"
/note="variant B"
1. 8793
/rpt_type=direct
59. 353
/note="heterogeneous telomeric unit"
rpt_type=tandem
/rpt_unit=tasyc
583. 3988
/gene="DR1"
/note="US22 gene family"
join(583. 841,955. 2975)
/gene="DR1"
/codon_start=1
/product="DR1"
/protein_id="AAD49614.1"
/db_xref="GI:5733511"

/translation="MPLARACHTLRLPLSYWMLLGRHSRLRHYSYLRKGLRL
PLPWEQECILHLPRKPKFLRYPCILTRQPHLLQGMADSSLMFDPKPYHPBSADSKL
PLGLITLSAFSKMVSSEPTKSGFHAADPLSLMTSGSPWLVLLQAPGSLFCHDVFOG
RLYLISHVSLFKTKGRLROCEAIYRAPLRLRVPLDPSIMTCRDPDAFLPKLRLAR
GLAATFALYRLHLHSGRSELSHVLEMERELVLTDRRGMPCTHLISGSEQSVSSD
AGTWNATKAKAGKEAEERGGROATRLASPHLTRRRCGSLGEEPSNADEF
ARCRPLDELCEGGLWLPFLTASPHVLLITEGPTVALDLNDTSLMRIDLELL
LRGLSILLSGRLPLRPPSGGEAARKPGYKEEGGRATTASATLSAPRPTPR
GVTEKRGVTGVPFSAHPESEEDTDGHHQDSGDDRGDGRGDRNDGARAND
ETPEQGRGHEDEORTDSREEDACEVARAREKTEGSGSCGATQYGGGE
HGWSSITLPSYRPPRPPVWPHLLFPLSPITVEDEPSARPCPGPAEPEK
SPPCPSPPADPSAVPRLSALSVSPSTARVPRFSLSSSSSSSSSSSSPSPL
SPSPSPSSPRSPFISPLRSPRLAKPWSGCHPAFPAPBSAIPPSKRVSPVSS
ASPSAPCIGRSRPPSAQTA"
complement(2723. 3325)
/gene="DR3"
/evidence=not_experimental
complement(2723. 3325)
/gene="DR3"
/codon_start=1
/product="DR3"
/protein_id="AAD49618.1"
/db_xref="GI:5733515"
/translation="MSRVSCVLACVACAGLCCMGVTCGDCOMRRRCARMGV
GPRVLDSGAMRVRSBGDAKSVSRCTETERPAPSAKSPYSPPVIVSSSSSSSCSAC
SSRVPSPSPPHASAVCAVSGKDLPHGADGDADGTDTLLEKGADEGAGNAT
GCPDTHGFARSPDMLGEMNGDLDGDEGTGEGDNGEGE"
3022. 3988
/gene="B1"
/evidence=not_experimental
3022. 3501
/gene="B1"
/codon_start=1
/product="B1"
/protein_id="AAD49617.1"
/db_xref="GI:5733514"
/translation="MKNMKTATKTKKRGKNGNPEETERMEPARSRTSAIPGLRR
SGPSTPLPGPEVRAHAPWTASATITDSDHSRISPTPSSGRGRIHRTGARTPRFS
AAEINGVYARAVALTKRKSETIDRLSLVPLGHGPHASLRSHLSALSALRPPDPK
3536. 3988
/gene="B2"
/evidence=not_experimental
3536. 3775
/gene="B2"
/codon_start=1
/product="B2"
/protein_id="AAD49616.1"
/db_xref="GI:5733513"
/translation="MQPMTKKKHTASKRSPARPSLSPLESRRRVGKSPERLHG
APPPRSPPLPSPSPSEHTRLSPORRRPQORKL"
3912. 3917
/note="DR1; B1; B2"
/evidence=not_experimental
3938. 3942
/note="DR1; B1; B2"
/evidence=not_experimental
3968. 3974
/note="DR1; B1; B2"
/evidence=not_experimental
3983. 3988
/note="DR1; B1; B2"
/evidence=not_experimental
5027. 7435
/gene="DR6"
/note="US22 gene family; transactivator"
join(5027. 5330,6329. 7203)
/gene="DR6"
/codon_start=1
/product="DR6"
/protein_id="AAD49615.1"
/db_xref="GI:5733512"
/translation="MTRRHTORTDGRGRIARNDGARLARARAREWILLARGPSKL
YGTSHRGERIHLPPWRWYCLELHPDYPDARSATVWGHMWGMPPTHVRRSVODCA

THIS PAGE BLANK (USPTO)

Query Match	4.48;	Score 101.4;	DB 72;	Length 162114;
Best Local Similarity	54.08;	Pred. No. 6.7e-07;		
Matches 207; Conservative	0;	Mismatches 176;	Indels 0;	Gaps 0

RESULT	13				
AF015297					
LOCUS	AF015297	4742 bp	mRNA	VRL	27-AUG-1997
DEFINITION	Human herpesvirus 6 (strain Uganda-1102)	IEZhom mRNA, complete	cds		
ACCESSION	AF015297				
VERSION	AF015297.1	GI:2343282			
KEYWORDS					
SOURCE					
ORGANISM					
	Human herpesvirus 6 (strain Uganda-1102).				
	Human herpesvirus 6 (strain Uganda-1102).				
	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;				
	Betaherpesvirinae; Roseolovirus.				

REFERENCE	JOURNAL	AUTHORS	TITLE	FEATURES
1 (bases 1 to 4742)	Soergel, A., Schlieve, U., Fleckenstein, B. and Neipel, F.	Identification of an human herpesvirus-6 transcription unit homologous to the immediate-early 2 gene of human cytomegalovirus - evidence for early/late gene unpublished	2 (bases 1 to 4742)	
Soergel, A., Schlieve, U., Fleckenstein, B. and Neipel, F.	Direct Submission	Submitted (21-Jul-1997)	Institut fuer Klinische und Molekulare Virologie, Universitaet Erlangen-Nuernberg, Schlossgarten 4, Erlangen D-91054, Germany	
location/Qualifiers				
1..4742				
/organism="Human herpesvirus 6 (strain Uganda-1102)"				
/db_xref="taxon:10370"				
1..4742				
/gene="ie2hom"				
196..4698				
/gene="ie2hom"				
/note="similar to the immediate-early 2 protein of human cytomegalovirus"				
/codon_start=1				
/product="ie2hom"				
/protein_id="AA867722.1"				
/db_xref="gi:2343283"				
/translation="MEPAKPGCGNMGSNDERMODYRPPDPMWESIKELIEELSLMCDTS /FDDLIATLESFGILIPESNNISNNVEBSDEKLTIAEKKOGKNDNDVDSAKK LSGLYCADIDHTQPLSDNTHQDPIYSQESIRFTKIQDPRIVAQTRQCTSSANQ SNEGSTQVRFASBLPNQLQPMYTSNNQANLNQNTFS.LPYQYRDPYRQIESIRE SRNNRGVDYFRRHHPYRRCGNGCKYNNYSQOYPKCFRTYRNRGRHSYD CSDSADLPHEHYTPYNEQONPPRMNMYKDFOLNKFENFESYDYSAFSDSTHVC OSNDNHPETKQOTIPETTTKTKHEATNDENSTNOVYLTPVILSYRPSYKMDLIK KIIDYDVLPELKEALTANGSRNDYDIOKKYKAHRCOSYVOKKRRSSQTNKDHNAS SRSDLEKRSNENEDKAVTKARDSKLPLSP.LPIPELADIPDADHTDKFYSPENR QIOUNLHRSKTSLODVPYISKHTPRAPTKDMSYKHHDSKNYRKMHSRPTSSKNN TNSNGDNKFEVSKYKNGSKATSTSPKSHYSSSDEGQYKSPVKTIQSPSYQ CKIKMPSIMDNKSAKNHNSADKNTLDSPI.RSNLNPFAEFKSNKSNITLSTSDSD CTDRPNCNSTRKSDKSDPNTKCGNSDKHLSKOTMAKSPASASASASASASAS SRASRASRSRDSRASRSRASRSRASRSRASRSRASRSRASRSRASRSRASRS ASKRSASRSRDSRASRSRASRSRASRSRASRSRASRSRASRSRASRSRASRS SRDSRASRSRDSRASRSRASRSRASRSRASRSRASRSRASRSRASRSRASRS ASSKRSASRSRNSRASRSRASRSRASRSRASRSRASRSRASRSRASRSRASRS SRASRELROYICDSNRKOTPPHDSINTKFEIIEIKRCEDLNFYKTKARLQCFNR HNDQYNNRPRPHITNTNRKSGESTNDDSESMRCKSGHCSNPSDILVYRKHKSGG SSSISSTIEBNCRSNHIYVTEKEKPTPYOSTRSSSSSSSSA.LSCSTLKTC RKTQYKDNKQIKSKSDSKHKTIDMSDYESNRHADYFKNSPDAEKPLHNSPENHR EOSNSENNAIDEOBKAPNITTSHLQGOQNRVLAHTKCKKRRPDDSDSINKFC KRTISAGKTESEVSDDLICYRDYRLERKVESEKFIHRGRVATKDFOKLEKNTMR FYKIOKPKPCNEKULKEAVYDICNGLSNNAALIMYTRSRKVAQILIKMOKELMTN PWITSEAFPRKNHAPKRYYDDEIKRFTIOLQKOGQELMDFENNTHDLSRSADV TNITATYRIDFVGAVTKCNKAKNPKEIYLRNIIIDGNPISITVPISEFKSKK STLSK"				
BASE COUNT	1772 a	1134 c	793 g	1043 t
ORIGIN				
Query Match	4.0%;	Score 92.2;	DB 72;	Length 4742;
Best Local Similarity	52.8%;	Pred. No. 6.9e-05;		
Matches 199;	Conservative 0;	Mismatches 178;	Indels 0;	Gaps 0.
QY	1561	cccaagaagtgtaacctgtgagagccatcggaaccccaagccagccagccaggaagact	1620	
DB	2462	CCAGCAAGGCTCCAGCAGGAGGATTTTCAGCAGACCTCCAGTAGAGCTCCAGAGGCGCT	2521	
QY	1621	gcagggagcagcagccgagagccatcggaagaccagagcccccgcgcgcgaggaacct	1680	
DB	2522	CAAGCAAGGCTCCAGAGAGGCTCCAGCAGACCTCCAGACAGAGGCTCCAGAGGAGATT	2581	
QY	1681	gcagggagcagcagccagcgcagagccatcggaaccccaagcccccgcgcgcgaggaacct	1740	
DB	2582	CTAGCAGGCGCTCCAGCAAGGCTCCAGCAGAGGCTCCAGCAGGATTTTCAGCAGAGCT	2641	

THIS PAGE BLANK (USPTO)

RESULT 12
AF157706/c 162114 DNA VRL 20-SEP-1999

LOCUS AF157706 Human herpesvirus 6B strain 229, complete genome.
DEFINITION AF157706 L13162 L14772 L16947
ACCESSION AF157706.1 GI:5733510
VERSION
KEYWORDS
SOURCE Human herpesvirus 6B.
ORGANISM Human herpesvirus 6B.
VIRUSES: dsDNA viruses, no RNA stage; Herpesviridae;
Betaherpesvirinae; Koseclovirus.
REFERENCE 1 (bases 19696 to 22886)
AUTHORS Pellett,P.E., Sanchez-Martinez,D., Dominguez,G., Black,J.B.,
Anton,E., Greenmoyer,C. and Dambaugh,T.R.
TITLE A strongly immunoreactive virion protein of human herpesvirus 6
variant B strain 229: Identification and characterization of the
gene and mapping of a variant-specific monoclonal antibody reactive
epitope
JOURNAL Virology 195 (2), 521-531 (1993)
MEDLINE 93331710
REFERENCE 2 (bases 64106 to 84963)
AUTHORS Stamey,F.R., Dominguez,G., Black,J.B., Dambaugh,T.R. and
Pellett,P.E.
TITLE Intragenomic linear amplification of human herpesvirus 6B orlyt
suggests acquisition of orlyt by transposition
J. Virol. 69 (1), 589-596 (1995)
MEDLINE 95074921
REFERENCE 3 (bases 64106 to 84963; 106012 to 126166)
AUTHORS Lindquester,G.J., Inoue,N., Allen,R.D., Castell,J.W., Stamey,F.R.,
Dambaugh,T.R., O'Brian,J.J., Danovich,R.M., Frenkel,N. and
Pellett,P.E.
TITLE Restriction endonuclease mapping and molecular cloning of the human
herpesvirus 6 variant B strain 229 genome
Arch. Virol. 141 (2), 367-379 (1996)
MEDLINE 96195263
REFERENCE 4 (bases 64106 to 84963)
AUTHORS Lindquester,G.J., O'Brian,J.J., Anton,E.D., Greenmoyer,C.A.,
Pellett,P.E. and Dambaugh,T.R.
TITLE Genetic content of a 20.9 kb segment of human herpesvirus 6B strain
229 spanning the homologs of human herpesvirus 6A genes U40-57 and
containing the origin of DNA replication
Arch. Virol. 142 (1), 103-123 (1997)
MEDLINE 97300856
REFERENCE 5 (bases 106012 to 126166)
AUTHORS Lindquester,G.J., Greenmoyer,C.A., Anton,E.D., O'Brian,J.J.,
Pellett,P.E. and Dambaugh,T.R.
TITLE Comparison of a 20 kb region of human herpesvirus 6B with other
human beta herpesviruses reveals conserved replication genes and
adjacent divergent open reading frames
Arch. Virol. 142 (1), 193-204 (1997)
MEDLINE 97300864
REFERENCE 6 (bases 1 to 162114)
AUTHORS Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S., Inoue,N. and
Pellett,P.E.
TITLE Human herpesvirus 6B genome sequence: coding content and comparison
with human herpesvirus 6A
J. Virol. 73 (10), 8040-8052 (1999)
MEDLINE 99412318
REFERENCE 7 (bases 19696 to 22886)
AUTHORS Pellett,P.E.
TITLE Direct Submission
Submitted (26-JUN-1993) Herpesvirus Section, Centers for Disease
Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta,
GA 30333, USA
REFERENCE 8 (bases 64106 to 84963; 106012 to 126166)
AUTHORS Pellett,P.E.
TITLE Direct Submission
Submitted (02-OCT-1993) Herpesvirus Section, Centers for Disease
Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta,
GA 30333, USA
REFERENCE 9 (bases 1 to 162114)
AUTHORS Pellett,P.E., Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S.
and Inoue,N.

TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Herpesvirus Section, Centers for Disease
Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta,
GA 30333, USA
REMARK Sequence updated by submitter
COMMENT On Aug 16, 1999 this sequence version replaced gi:405170 gi:405156
gi:305397.

FEATURES
source Location/Qualifiers
1..162114
/organism="Human herpesvirus 6B"
/strain="229"
/db_xref="taxon:32604"
/note="variant B"
1..8793
/rpt_type-direct
59..353
/note="heterogeneous telomeric unit"
/rpt_type-tandem
583..3988
/rpt_unit-tarsyc
/gene="DRI"
/note="GS22 gene family"
join(583..841,955..2975)
/gene="DRI"
/codon_start=1
/product="DRI"
/protein_id="A049614.1"
/db_xref="GI:5733511"
/translacion="MPLTARAGHTLRLPLSHYWWLLGRHSRLHVSHTLRRLKRL
PLPWEQECILHHPKRYKFLRPLCTROPHLIQGMPADSSLMFPKPYHPADSKL
PGLITLTSFMSRVSSEPHNGSGFHAADSLSWLTSSPMVLVLAQAPGSLCHDFVG
KYLILSHSVSLFKTGLRQCEALYRAPIARVPLPSIMTCRDPDAPLILARTAR
GLAARVLRHLGSRSELHSHVLRMERELVLTTRRGMPCTHLLSGESQVSSD
ACDTPVAATEKAGAKKEAEKRGQADRLAPHLTRGRDCCSLQCEESAEDE
ACRPLDELCEGWLPAFLYASPHVCLITBEGPVLAIDLNTSLMRIDDEL
LRGSLILSLGLRLPLRPSGSGEAKRGYKEGSGRATATATISPRPTPR
GYTERGRVITGVPEPSAHPESEOTDHHGROSGHGDGGRGDRDGRAND
ETRPQORGHDEGEOTDNGREDDQSEVARRDEGTBOGSGRGRATQYGGRG
HGAWSTPLSVAPDPKRVVPPHLLFPSPILSTYVPEPSARRCPGPAEESK
SPQPCPSPLASQAVPRLSAUSVPSSTARKYRFSLSLSSSSSSSSSSSSSSSPS
SPSPVPSPPSP
ASPSAPCIGRSRPPSAQTA"
/complement(2723..3325)
/gene="DR3"
/evidence-not_experimental
/complement(2723..3325)
/gene="DR3"
/codon_start=1
/product="DR3"
/protein_id="A049618.1"
/db_xref="GI:5733515"
/translacion="MSRVFSCVLRACVACAGLCWCVMGVCDCORWRRRCARRY
GPRVLDGAMRVSGDARSVSTCTETAPSPARSPVSPPLVSSSSSSSSSAC
SSRVPSPPSPHAASHAVCAEGGRDLPMHAGADADADGDTLLERGADEGAGNAT
GCPEDTHGARSPGDLMEWNGDLCEGETGEGDNGEGE"
3022..3988
/gene="B1"
/evidence-not_experimental
3022..3501
/gene="B1"
/codon_start=1
/product="B1"
/protein_id="A049617.1"
/db_xref="GI:5733514"
/translacion="MKNMKTATKTKRGRKRGKNTPTERRRREPASTAIPSGLAR
SGESTPLRPGPEVHAAPTATASADSHRISPPVPSRGRHTRGARTRES
AAEINGYARAVTRKRTSETIDRLSLVPLGHPHSLRSLRLARALRPPDP
3536..3988
/gene="B2"
/evidence-not_experimental
3536..3775
/gene="B2"
/codon_start=1

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

```
/product="Vmw63(ICP27)"
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

6512. 7150
/gene="DR7"
/function="transforming activity"
/note="42.2% identical to DR7 gene of strain U1102 of
HHV-6
US22 gene family"
/codon_start=1
/product="transactivator"
/protein_id="BA78217.1"
/db_xref="GI:4995984"
/translation="MRAALCRASNTVFARGLQRPVAYEEDLELDRMGSTNV
RLFFVIAAADDEHDLTVGGICOTHCAGSCLETPVDPPTAAGVMTITVAQOY
EYETAMVIVYDRCPTASGPITASAVWRTSPPCGMAAGHETLDVFGPC"
complement(7236..7706)
/gene="DRH2"
complement(7236..7706)
/note="ORF unique to HST strain"
/codon_start=1
/protein_id="BA78218.1"
/db_xref="GI:4995985"
/translation="MLAPTPRGVYIKGCGALSSPDLCSPCVAGIGRCYSNGARG
RRLDIDRYVFLGCVSNTERRKIVCCVHSPIELIGLKRGDTVRYTHPPTTR
7928. 8662
/gene="DR8"
7928. 8662
/note="DR8"
/note="13.6% identical to DR8 gene of strain U1102 of
HHV-6"
/codon_start=1
/protein_id="BA78219.1"
/db_xref="GI:4995986"
/translation="NRQTVTRDKGRLLANOLGKKNRCKILTKETSLHLPST
TTPPLRNHCPRESRHVOTQTRHATTSMAVAGQAGSIRGADVSMNGIARQY
PSSRTIALRQCKPLASISITLTLTLTLTLTLTLTLTLTLTLTLTLTLTLTL
LTLTLTLTLTLHPPTARAPLMEACFPKAPLAEAKISPPORAVYVACRR
CAPRGSPLTPPRNGRGYGGKRVY"
complement(8292..8807)
/gene="LJ1"
complement(8292..8807)
/note="LJ1"
/note="16.6% identical to L1 gene of strain U1102 of
HHV-6"
/codon_start=1
/protein_id="BA78220.1"
/db_xref="GI:4995987"
/translation="MALAAYLFLRLLVCPASALPSASLISRAFPSPSHRAPAG
CTFAFKLHAIRHPLARSISGGGGGIGAGAGTCTTRGLGVREGRTFAPLAVARV
VKNAPRIEAAARALGDLGDLGDLGDLGDLGDLGDLGDLGDLGDLGDLGDLG
LGLGDLGDLG"
8929. 9384
/gene="U1"
8929. 9384
/note="U1"
/note="25.6% identical to U1 gene of strain U1102 of
HHV-6"
/codon_start=1
/protein_id="BA78221.1"
/db_xref="GI:4995988"
/translation="MRAPRRALDLSDDPPISSVMOJRTTRTDMPTRPADCHVSE
SIHHEKIHTRKTOENTHOKTHKIHOKTHKIHOKTHKIHOKTHKIHOKTHKIHOK
HIKVIHTKIVYTHKTHKIHKTHKIHKTHKIHKTHKIHKTHKIHKTHKIHKTHK
complement(9467..10768)
/gene="U2"
complement(9467..10768)
/note="U2"
complement(9467..10768)
/note="92% identical to U2 of strain U1102 of HHV-6
US22 gene family"
/codon_start=1
/protein_id="BA78222.1"
/db_xref="GI:4995989"

Query Match 4.5%: Score 103.2; DS 72; Length 161573;
Best Local Similarity 54.8%; Pred. No. 3,4e-07;
Matches 204: Conservative 0; Mismatches 168; Indels 0; Gaps 0;
CDS
gene
/translation="MCSQFCGRSVYHEFRGRREYVDSITLTVENSCLVNIIVYVR
LPVLPSPFFFFILLISRIPIHKKKFCSPFISISMSLAAALCPSSCSFSGRD
LRDDAAVEYKMTAFVTVARPLSADVDVSTPACHTALSMVPGALRATISM
IGITQALATLSRGYCCGPGMDLVITVITFAEYSALVLCAGEYFNGVDDALY
RLADBAFLKMGHGRREPYVGSCLMETASFEGSGVDALAAVAFDALKVLP
WPAAGFEFAVPRRAEKRMRLIPGGGAAGVIGFFGCGVTLPLRROVLMDOVGRV
AASIDGAVVRLSDSFRAFLAMVKRLFKHNRFPGLMTMOLPYCVHAPVNLPAV
YOLSPHVEREMSAVSGASTVVRDCEDTLRDGDAGVDIS"
complement(10891..12051)
/gene="U3"
complement(10891..12051)
RESULT 12
AF157706/c
LOCUS 162114 bp DNA VRL 20-SEP-1999
DEFINITION Human herpesvirus 6B strain 229, complete genome.
ACCESSION AF157706 L13162 L14772 L16947
VERSION AF157706.1 GI:5733510
KEYWORDS
SOURCE human herpesvirus 6B.
ORGANISM Human herpesvirus 6B.
VIRUSES; dsDNA viruses, no RNA stage; Herpesviridae;
Betaherpesvirinae; Roseolovirus.
1 (bases 19696 to 22886)
REFERENCE
AUTHORS Pellett,P.E., Sanchez-Martinez,D., Dominguez,G., Black,J.B.,
Anton,E., Greenmeyer,C. and Dambaugh,T.R.
A strongly immunoreactive virion protein of human herpesvirus 6
variant B strain 229: identification and characterization of the
gene and mapping of a variant-specific monoclonal antibody reactive
epitope
JOURNAL Virology 195 (2), 521-531 (1993)
MEDLINE 93331710
REFERENCE 2 (bases 64106 to 84963)
AUTHORS Stamey,F.R., Dominguez,G., Black,J.B., Dambaugh,T.R. and
Pellett,P.E.
TITLE Intragenomic linear amplification of human herpesvirus 6B orllyt
suggests acquisition of orllyt by transposition
JOURNAL J. Virol. 69 (1), 589-596 (1995)
MEDLINE 95074921
REFERENCE 3 (bases 64106 to 84963; 106012 to 126166)

THIS PAGE BLANK (USPTO)

[illegible]

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

/ gene- "KL2"

THIS PAGE BLANK (USPTO)

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 17:09:44 ; Search time 262.03 Seconds
(Without alignments)
1389.877 Million cell updates/sec

Title: US-09-431-843b-5
Perfect score: 2408
Sequence: 1 tagaattcagcgcgcgtga.....ccctctgtgagtg99999cgcg 2408

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5C_COMB.seq:*
5: /cgn2_6/prodata/2/ina/5D_COMB.seq:*
6: /cgn2_6/prodata/2/ina/5E_COMB.seq:*
7: /cgn2_6/prodata/2/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	4.2	6192	3	US-08-479-537A-1
2	101	4.2	6449	3	US-08-479-537A-4
3	76	3.2	7218	1	US-08-332-463-14
4	74.6	3.1	1931	4	US-09-130-114-2
5	68.2	2.8	35100	3	US-08-770-379-17
6	67.2	2.8	2580	5	US-09-050-863-2
7	67.2	2.8	5452	4	US-09-130-114-1
8	67.2	2.8	10596	1	US-07-884-811-15
9	67.2	2.8	10596	1	US-07-885-971-15
10	67.2	2.8	10596	1	US-08-197-783A-15
11	67.2	2.8	10596	1	US-08-194-088B-15
12	67.2	2.8	10596	1	US-08-194-087-15
13	67.2	2.8	10596	6	PCT-US93-04648-15
14	61.8	2.6	8438	1	US-07-945-283-1
15	60.2	2.5	3489	4	US-08-728-323A-1
16	60.2	2.5	32207	3	US-08-770-379-20
17	58.8	2.4	711	2	US-08-452-531-2
18	58.8	2.4	711	2	US-08-460-746A-2
19	58.8	2.4	711	2	US-08-460-555-2
20	58.8	2.4	1091	2	US-08-452-531-1
21	58.8	2.4	1091	2	US-08-460-746A-1
22	58.8	2.4	1091	2	US-08-460-555-1
23	58.4	2.4	150	3	US-07-829-461A-8
24	58.4	2.4	51259	5	US-08-781-891-209
25	56.2	2.3	2277	2	US-08-676-967-5
26	56.2	2.3	2277	2	US-08-676-974-5

27	56.2	2.3	2277	4	US-09-098-487-5	Sequence 5, Appl 1
28	56	2.3	16442	5	US-08-781-891-208	Sequence 208, App
29	54.8	2.3	768	5	US-08-946-026-32	Sequence 32, Appl
30	53	2.2	2824	4	US-09-010-928B-3	Sequence 3, Appl 1
31	53	2.2	6192	3	US-08-479-537A-1	Sequence 1, Appl 1
32	53	2.2	6449	3	US-08-479-537A-4	Sequence 4, Appl 1
33	50.8	2.1	2277	2	US-08-676-967-2	Sequence 2, Appl 1
34	50.8	2.1	2277	2	US-08-676-974-2	Sequence 2, Appl 1
35	50.8	2.1	2277	2	US-09-098-487-2	Sequence 2, Appl 1
36	50.8	2.1	2301	2	US-08-306-691B-23	Sequence 23, Appl 1
37	50.8	2.1	2301	6	PCT-US93-06251-78	Sequence 78, Appl 1
38	50.6	2.1	53526	5	US-08-658-136-2	Sequence 2, Appl 1
39	50.6	2.1	53577	5	US-08-658-136-1	Sequence 1, Appl 1
40	48	2.0	1052	2	US-08-466-603-1	Sequence 1, Appl 1
41	48	2.0	1052	2	US-08-314-503A-1	Sequence 1, Appl 1
42	48	2.0	1052	2	US-08-468-066-1	Sequence 1, Appl 1
43	48	2.0	1052	3	US-08-466-717-1	Sequence 1, Appl 1
44	48	2.0	1052	5	US-08-466-743-1	Sequence 1, Appl 1
45	48	2.0	1052	6	PCT-US95-12414-1	Sequence 1, Appl 1

ALIGNMENTS

Sequence Comparison H

RESULT 1
US-08-479-537A-1
Sequence 1, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

THIS PAGE BLANK (USPTO)

```

LENGTH: 16192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note="The nucleotides spanning
OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note="Nucleotide 457 is X1 - NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC, CCG, CCA
OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note="Nucleotide 487 is Y - NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn - AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note="Nucleotide 496 is X2 - NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, GCC, CCA
OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."

```

Query Match 4.2%; Score 101; DB 3; Length 6192;
Best Local Similarity 45.4%; Pred. No. 2e-12;
Matches 215; Conservative 0; Mismatches 255; Indels 0; Gaps 0.

```

RESULT 2
US-08-479-537A-4
: Sequence 4, Application US/08479537A
: Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning
OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat 1
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fl
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..5661
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 17:20:35 ; Search time 262.03 Seconds

(Without alignments)
1355.245 Million cell updates/sec

Title: US-09-431-843b-9
Perfect score: 2348
Sequence: 1 tagaattcagcgcgcctga.....ccctctgagtgaggagcg 2348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Maximum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5C.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/6.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/PCRUS.COMB.seq:*
7: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.4	3.8	6192	US-08-479-537A-1	Sequence 1, Appl1
2	88.4	3.8	6449	US-08-479-537A-4	Sequence 4, Appl1
3	76	3.2	7218	US-08-232-463-14	Sequence 14, Appl1
4	74.6	3.2	1931	US-09-130-114-2	Sequence 2, Appl1
5	70.2	3.0	35100	US-08-770-379-17	Sequence 17, Appl1
6	67.2	2.9	2580	US-08-050-863-2	Sequence 2, Appl1
7	67.2	2.9	5452	US-09-130-114-1	Sequence 1, Appl1
8	67.2	2.9	10596	US-07-884-811-15	Sequence 15, Appl1
9	67.2	2.9	10596	US-07-885-971-15	Sequence 15, Appl1
10	67.2	2.9	10596	US-08-087-783A-15	Sequence 15, Appl1
11	67.2	2.9	10596	US-08-194-088B-15	Sequence 15, Appl1
12	67.2	2.9	10596	US-08-194-087-15	Sequence 15, Appl1
13	67.2	2.9	10596	US-07-945-283-1	Sequence 15, Appl1
14	61.8	2.6	8438	US-08-728-323A-1	Sequence 1, Appl1
15	60.2	2.6	3489	US-08-770-379-20	Sequence 20, Appl1
16	60.2	2.6	32207	US-08-770-379-20	Sequence 20, Appl1
17	58.4	2.5	150	US-07-829-461A-8	Sequence 8, Appl1
18	58.4	2.5	51259	US-08-781-891-209	Sequence 209, App
19	56.2	2.4	2277	US-08-676-967-5	Sequence 5, Appl1
20	56.2	2.4	2277	US-08-676-974-5	Sequence 5, Appl1
21	56.2	2.4	2277	US-09-098-487-5	Sequence 5, Appl1
22	56	2.4	16442	US-08-781-891-308	Sequence 208, App
23	54.8	2.3	768	US-08-946-026-32	Sequence 32, Appl1
24	53.2	2.3	711	US-08-452-531-2	Sequence 2, Appl1
25	53.2	2.3	711	US-08-460-76A-2	Sequence 2, Appl1
26	53.2	2.3	711	US-08-460-555-2	Sequence 2, Appl1

C 27	53.2	2.3	1091	2	US-08-452-531-1	Sequence 1, Appl1
C 28	53.2	2.3	1091	2	US-08-460-746A-1	Sequence 1, Appl1
C 29	53.2	2.3	1091	2	US-08-460-555-1	Sequence 1, Appl1
C 30	52.6	2.2	2824	4	US-09-010-928B-3	Sequence 3, Appl1
C 31	51.4	2.2	2830	4	US-09-010-928B-1	Sequence 1, Appl1
C 32	50.8	2.2	2277	2	US-08-676-967-2	Sequence 2, Appl1
C 33	50.8	2.2	2277	2	US-08-676-974-2	Sequence 2, Appl1
C 34	50.8	2.2	2277	4	US-09-098-487-2	Sequence 2, Appl1
C 35	50.8	2.2	2301	2	US-08-306-691B-23	Sequence 23, Appl1
C 36	50.8	2.2	2301	6	PCT-US93-06251-78	Sequence 78, Appl1
C 37	50.6	2.2	53526	5	US-08-658-136-2	Sequence 2, Appl1
C 38	50.6	2.2	53577	5	US-08-658-136-1	Sequence 1, Appl1
C 39	49.4	2.1	2830	4	US-09-010-928B-1	Sequence 1, Appl1
C 40	48.2	2.1	15202	5	US-08-922-635-21	Sequence 21, Appl1
C 41	48	2.0	1052	2	US-08-466-603-1	Sequence 1, Appl1
C 42	48	2.0	1052	2	US-08-314-503A-1	Sequence 1, Appl1
C 43	48	2.0	1052	2	US-08-468-066-1	Sequence 1, Appl1
C 44	48	2.0	1052	3	US-08-466-717-1	Sequence 1, Appl1
C 45	48	2.0	1052	6	PCT-US93-12414-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-479-537A-1
Sequence 1, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMBER, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREVEN, Maïa

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

THIS PAGE BLANK (USPTO)

Query Match	3.8%	Score 88.4;	DB 3;	Length 6192;
Best Local Similarity	45.4%	Pred. No. 1.3e-09;		
Matches 188; Conservative	0;	Mismatches 226;	Indels 0;	Gaps 0;

```

US-08-479-537A-4
Sequence 4, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning
OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..5661
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 - NN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - Cct, Ccc,
OTHER INFORMATION: or Ccg; and Ala - Gct, Gcc, Gca, or Gcg."
FEATURE:

```

THIS PAGE BLANK (USPTO)

Query Match 3.3%; Score 76; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred. No. 4.3e-07;
Matches 25; Conservative 240; Mismatches 155; Indels 0; Gaps 0;

942 gctcgaagagctccgaaggttgagaggaagagcccgagagcccgagagc 1001
1452 GATGAGGAATTGGTACRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1393
1002 cagcaccagagctcgagcctgtgagccagacagagaggtgagagcga 1061
1392 RRR 1333
1062 ggggccccagcagagcgctgagcccgagatcgagacccctgagagagcag 1121
1332 RRR 1273
1122 ggaatgagcagagcgagcagagatagcagagcccttaagcccaagagagca 1181
1272 RRR 1213
1182 gaagagagagctgagcctgagccgagagcagagcccgagagagcag 1241
1212 RRR 1153
1242 tccctcagagcgagagagatcgctcgaatttgagaggtgtgctcctcagcagagcag 1301
1152 RRR 1093
1302 cctcagagagagcagagagagtgagcagcagcctgagagcagcagcctcagcctc 1361
1092 RRR 1033

Sequence Unpaired

RESULT 2
US-08-479-537A-1
Sequence 1, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMBER, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: HAREVENI, Maïa
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat 1
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fl
OTHER INFORMATION: The number of such repeats varies from 1 to 80."

FEATURE:
NAME/KEY: repeat_region
LOCATION: 121..6166
OTHER INFORMATION: /note= "Nucleotide 457 is X1 - NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC,
OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."

FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y - NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC,
OTHER INFORMATION: or ACG; and Asn - AAT or AAC."

FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 - NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC,
OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."

Query Match 3.3%; Score 75.8; DB 3; Length 6192;
Best Local Similarity 45.5%; Pred. No. 4.6e-07;
Matches 161; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

1585 ccacgagagagcccgagcccgagcccgagagcctgagagcagcagcagagc 1644
461 CCCACGGGTGTCACCTCGCGCCCGGACNNNAGGCCGNNCCGGGCTCCACCGCCCCNNNG 520
1645 ccacgagagagcccgagcccgagcccgagcctgagagcagcagcagcagcagc 1704
521 CCCACGGGTGTCACCTCGCGCCCGGACNNNAGGCCGNNCCGGGCTCCACCGCCCCNNNG 580
1705 ccacgagagagcccgagcccgagcccgagcctgagagcagcagcagcagcagc 1764
581 CCCACGGGTGTCACCTCGCGCCCGGACNNNAGGCCGNNCCGGGCTCCACCGCCCCNNNG 640
1765 ccacgagagagcccgagcccgagcccgagcctgagagcagcagcagcagcagc 1824
641 CCCACGGGTGTCACCTCGCGCCCGGACNNNAGGCCGNNCCGGGCTCCACCGCCCCNNNG 700
1825 ccacgagagagcccgagcccgagcccgagcctgagagcagcagcagcagcagc 1884
701 CCCACGGGTGTCACCTCGCGCCCGGACNNNAGGCCGNNCCGGGCTCCACCGCCCCNNNG 760

THIS PAGE BLANK (USPTO)

LOCATIONS
FEATURES

```

sequence 2, 6268807
Patent No. 5976807
GENERAL INFORMATION:
APPLICANT: Horlick, Robert A.
APPLICANT: Damaaj, Baasam B.
APPLICANT: Robbins, Alan K.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
TITLE OF INVENTION: From Multiple Transfected Episomes
FILE REFERENCE: 0867/1D903051
CURRENT APPLICATION NUMBER: US/09/130.114
CURRENT FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1931
TYPE: DNA
ORGANISM: EBNA
US-09-130-114-2

```

Query Match	3.38; Score 74.6; DB 4; Length 1931;
Best Local Similarity	63.88; Pred. No. 6.4e-07;

THIS PAGE BLANK (USPTO)


```

Query Match Similarity      3.3%; Score 76; DB 1; Length 7218;
Best Local Similarity      6.0%; Pred. NO.3.9e-07;
Matches      25; Conservative 240; Mismatches 155; Indels      0; Gaps      0;

OY      942 gctcgaaggctccgaaggtgagaggaagaaagcccgaggaccaccagcagc 1001
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1452 GATAGACAATTGTGACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1393

OY      1002 cagcaccgccggctggacctgtggtccaaagcatagcaaggttgaggaggttgagca 1061
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1392 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1333

OY      1062 ggggccccacgacgagctgtgagccccagagatgcgggacccttgaaagagcagcg 1121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1332 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1273

OY      1122 ggatgagcgaaggcgcccaeggggaagatggtccggaagccctaaagcccaagagca 1181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1272 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213

OY      1182 gaagaggaagctcgagctgagccgagcgagcagccgcccacagagcagccctcagag 1241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1212 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1153

OY      1242 tgcctcagaagtgagggaagatgcctctgaattgaggggtgtgccctcagcagcgag 1301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1152 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1093

OY      1302 cctcgaagcgaggaccgaaggtggcggtcagacacctgtggaggaagtcagccctg 1361
      : : : : : : : : : : : : : : : : : : | | | | | | | | | |
DB      1092 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1033

RESULT      2
US-08-479-537A-1
: Sequence 1, Application US/08479537A
: Patent No. 5861381
: GENERAL INFORMATION:
: APPLICANT: CHAMON, Pierre
: APPLICANT: KIENV, Marie-Paule
: APPLICANT: LATHE, Richard
: APPLICANT: HAREUVENT, Mara
: TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/479,537A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 90/13101
: FILING DATE: 23-OCT-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/FR91/00835
: FILING DATE: 23-OCT-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/039,320
: FILING DATE: 04-APR-1993
: PRIOR APPLICATION DATA: US 08/403,576

```

```

? FILING DATE: 14-MAR-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Teskin, Robin L.
? REGISTRATION NUMBER: 35,030
? REFERENCE/DOCKET NUMBER: 017753-025
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 836-6620
? TELEFAX: (703) 836-2021
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 6192 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: sig_peptide
? LOCATION: 58..120
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 439..5239
? OTHER INFORMATION: /note= "The nucleotides spanning
? OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat i
? OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fl
? OTHER INFORMATION: The number of such repeats varies from 1 to 80."
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 121..6166
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 457
? OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
? OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
? OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 487
? OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
? OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
? OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 496
? OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
? OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
? OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-1

Query Match 3.3%; Score 75.8; DB 3; Length 6192;
Best Local Similarity 45.5%; Pred. No. 4.1e-07;
Matches 161; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 1585 ccatcgagagaccccaaggccccgcgaccagcaggaactgcagggagcagcagaagcagcagc 1644
Db 461 CCACAGGtGTACCTCGCCGCCGGACACNNNAGGCCGNNCCGGGCTCCACCgCCCCCnnNG 520
QY 1645 ccatcgagagaccccaaggccccgcgaccagcaggaactgcagggagcagcagaagcagcagc 1704
Db 521 CCCAGGtGTACCTCGCCGCCGGACACNNNAGGCCGNNCCGGGCTCCACCgCCCCCnnNG 580
QY 1705 ccatcgagagaccccaaggccccgcgaccagcaggaactcgaagtatgatgcagccagcagcagc 1764
Db 581 CCACAGGtGTACCTCGCCGCCGGACACNNNAGGCCGNNCCGGGCTCCACCgCCCCCnnNG 640
QY 1765 ccatcgagagaccccaaggccccgcgaccagcaggaactgcagggagcagcagaagcagcagcagc 1824
Db 641 CCCAGGtGTACCTCGCCGCCGGACACNNNAGGCCGNNCCGGGCTCCACCgCCCCCnnNG 700
QY 1825 ccatcgagagaccccaaggccccgcgaccagcaggaactgcagggagcagcagaagcagcagcagc 1884
Db 701 CCACAGGtGTACCTCGCCGCCGGACACNNNAGGCCGNNCCGGGCTCCACCgCCCCCnnNG 760

```

Qy 1885 ccacgcgagaccccccagcccccagcagcctacacagggatgagccagcc 1938
Db 761 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCC 814

RESULT 3

US-08-479-537A-4
Sequence 4, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENE, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 01753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning
OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..5661
FEATURE:

NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-4

Query Match 3.3%; Score 75.8; DB 3; Length 6449;
Best Local Similarity 45.3%; Pred. No. 4.1e-07;
Matches 161; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

Qy 1585 ccacgcgagaccccccagcccccagcagcagcctgcaaggagcagcagcagcagc 1644
Db 461 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCCCNNNG 520
Qy 1645 ccacgcgagaccccccagcccccagcagcagcctgcaaggagcagcagcagcagc 1704
Db 521 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCCCNNNG 580
Qy 1705 ccacgcgagaccccccagcagcagcagcagcagcctgcaaggagcagcagcagcagc 1764
Db 581 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCCCNNNG 640
Qy 1765 ccacgcgagaccccccagcagcagcagcagcagcagcagcctgcaaggagcagcagcagc 1824
Db 641 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCCCNNNG 700
Qy 1825 ccacgcgagaccccccagcagcagcagcagcagcctgcaaggagcagcagcagcagc 1884
Db 701 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCCCNNNG 760
Qy 1885 ccacgcgagaccccccagcagcagcagcagcagcctgcaaggagcagcagcagcagc 1938
Db 761 CCCACGGGTGTCACCTCGGCCCCGCGACNNNAGCGCCGNNCCGGGCTCCACCGCCCC 814

RESULT 4

US-09-130-114-2/c
Sequence 2, Application US/09130114
Patent No. 5976807
GENERAL INFORMATION:
APPLICANT: Horlick, Robert A.
APPLICANT: Robbins, Basam B.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
FROM MULTIPLE TRANSFECTED EPISOMES
FILE REFERENCE: 0867/1D903051
CURRENT APPLICATION NUMBER: US/09/130,114
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1931
TYPE: DNA
ORGANISM: EBNA
US-09-130-114-2

Query Match 3.3%; Score 74.6; DB 4; Length 1931;
Best Local Similarity 63.8%; Pred. No. 5.7e-07;

Matches 113; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 10 ggcgcctgaatctctgagcagatgagaccccgactgctacacttggaag 69
DB 1003 GTGACGAGAGAGCTGGGCGCGAGCTGAGAGACGAGGAGGAGGAGGAG 944

QY 70 gacagagagatgagagagagcagagagagactgagagagcagagcgcgcgc 129
DB 943 GAGACAGAGAGAGAGGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAG 884

QY 130 gacagagagcagcagcagcagcagcagcagcagcagcagcagcagcag 186
DB 883 GGGAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 827

RESULT 5
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 611411
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-09-050-863-2

Query Match 2.9%; Score 67.2; DB 5; Length 2580;
Best Local Similarity 47.7%; Pred. No. 1.9e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggtgag 1015
DB 707 GGCAGAGACAG 766

QY 1016 ggaactgtgagcagagcagagcagagagagagagagagagagagagag 1072
DB 767 AGGGCAG 826

QY 1073 cagcagagcgtgagagagagagagagagagagagagagagagagagag 1132

DB 827 AGAGCAG 886

QY 1133 ggggccaag 1192

DB 887 GGCAG 946

QY 1193 tggagctgagagcagcagcagcagcagcagcagcagcagcagcagcag 1252

DB 947 AGGGCAG 1006

QY 1253 tggagagagatcgcttgatatttgagagagagagagagagagagagag 1312

DB 1007 AGGGCAG 1066

QY 1313 ggaaccag 1372

DB 1067 AGGGCAG 1126

QY 1373 tggagagcagaggtgagcagcagcagcagcagcagcagcagcagcagc 1432

DB 1127 AGGGCAG 1186

RESULT 6
US-09-130-114-1/C
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Robert A.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903051
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
; US-09-130-114-1

Query Match 2.9%; Score 67.2; DB 4; Length 5452;
Best Local Similarity 47.7%; Pred. No. 2.2e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggtgag 1015

DB 2098 GGCAG 2039

QY 1016 ggaactgtgagcagagcagagcagagagagagagagagagagagagag 1072

DB 2038 AGGGCAG 1979

QY 1073 cagcagagcgtgagagagagagagagagagagagagagagagagagag 1132

DB 1978 AGGAGCAG 1919

QY 1133 ggggccaag 1192

DB 1318 GGCAG 1859

QY 1193 tggagctgagagcagcagcagcagcagcagcagcagcagcagcagcag 1252

DB 1858 AGGGCAG 1799

QY 1253 tggagagagatcgcttgatatttgagagagagagagagagagagagagag 1312

DB 1798 AGGGCAG 1739


```

1  APPLICANT:  Cheung, Andrew K.
2  APPLICANT:  Wesley, Ronald D.
3  TITLE OF INVENTION:  Pseudorabies Virus Deletion Mutants
4  TITLE OF INVENTION:  Involving The EP0 and LfT Genes
5  NUMBER OF SEQUENCES:  7
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE:  Curtiss P. Ribando
8  STREET:  1815 No. 5352596th University Street
9  CITY:  Peoria
10 STATE:  IL
11 COUNTRY:  USA
12 ZIP:  61604
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  Patentln Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/07/945,283
20 FILING DATE:  19920911
21 CLASSIFICATION:  424
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Ribando, Curtiss P
24 REGISTRATION NUMBER:  27976
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE:  309-685-4011 ext.513
27 TELEFAX:  309-685-4128
28 INFORMATION FOR SEQ ID NO:  1:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH:  8438 base pairs
31 TYPE:  NUCLEIC ACID
32 STRANDEDNESS:  double
33 TOPOLOGY:  linear
34 MOLECULE TYPE:  DNA (genomic)
35 HYPOTHEICAL:  NO
36 ANTI-SENSE:  NO
37 ORIGINAL SOURCE:
38 ORGANISM:  Pseudorabies virus
39 FEATURE:
40 NAME/KEY:  CDS
41 LOCATION:  622..6495
42 FEATURE:
43 NAME/KEY:  variation
44 LOCATION:  replace(1099, "g")
45 FEATURE:
46 NAME/KEY:  variation
47 LOCATION:  replace(1267, "t")
48 FEATURE:
49 NAME/KEY:  variation
50 LOCATION:  replace(1381, "c")
51 FEATURE:
52 NAME/KEY:  variation
53 LOCATION:  replace(1566, "c")
54 FEATURE:
55 NAME/KEY:  variation
56 LOCATION:  replace(7010, "g")
57
58 US-07-945-283-1

```

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

[illegible]

Mon Oct 16 10:08:46 2000

us-09-431-843b-4.rni

Page 10

Search completed: October 12, 2000, 17:09:44
Job time: 28148 sec

```

1 APPLICANT: Wesley, Ronald D.
2 TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
3 NUMBER OF INVENTION: Involving The EPO and LTR Genes
4 CORRESPONDENCE ADDRESS:
5 ADDRESS: Curtis P. Ribando
6 STREET: 1815 No. 5352596th University Street
7 CITY: Peoria
8 STATE: IL
9 COUNTRY: USA
10 ZIP: 61604
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent In Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/07/945,283
19 FILING DATE: 19920911
20 CLASSIFICATION: 424
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Ribando, Curtis P
23 REGISTRATION NUMBER: 27976
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 309-685-4011 ext. 513
26 TELEFAX: 309-685-4128
27 INFORMATION FOR SEQ ID NO: 1:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 8438 base pairs
30 TYPE: NUCLEIC ACID
31 STRANDEDNESS: double
32 TOPOLOGY: linear
33 MOLECULE TYPE: DNA (genomic)
34 HYPOTHEICAL: NO
35 ANTI-SENSE: NO
36 ORIGINAL SOURCE:
37 ORGANISM: Pseudorabies virus
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: 622..6495
41 FEATURE:
42 NAME/KEY: variation
43 LOCATION: replace(1099, "g")
44 FEATURE:
45 NAME/KEY: variation
46 LOCATION: replace(1267, "t")
47 FEATURE:
48 NAME/KEY: variation
49 LOCATION: replace(1381, "c")
50 FEATURE:
51 NAME/KEY: variation
52 LOCATION: replace(1566, "c")
53 FEATURE:
54 NAME/KEY: variation
55 LOCATION: replace(7010, "g")
56
57 US-07-945-283-1

```

[illegible]

RESULT 3
 US-08-781-891-209/c
 : Sequence 209, Application US/08781891
 Patent No. 6090620
 GENERAL INFORMATION:
 APPLICANT: Fu, Ying-Hui
 APPLICANT: Yu, Chang-En
 APPLICANT: Oshima, Junko
 APPLICANT: Mulligan, John T.
 APPLICANT: Schellenberg, Gerald D.
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 NUMBER OF SEQUENCES: 209
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 City: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,891
 FILING DATE: 27-DEC-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6090620tenburg Ph. D., Carol
 REGISTRATION NUMBER: 39,311
 REFERENCE/DOCKET NUMBER: 240052.419
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 209:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 51259 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-781-891-209

```

Query Match:          4.7% ; Score 58.4 ; DB 5; Length 51259;
Best Local Similarity 66.9%; Pred. No. 0.00054;
Matches      83; Conservative    0; Mismatches   41; Indels     0; Gaps     0.

QY      64 ggcggcgagcaggagatgctggagacgcggagagcagagactgtcgaagaccgcgaagc 123
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      251 GGAGGGGCGAGGAGGAGGAGAGGAAGAAGAGACAGAGAGAGAGAGAGAAAGCA 192

QY      124 cgcgcgcgcgaagggacgcgcggagccggggagacagagagtgctgagaagccgcgggc 183
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      191 GGAGGAGGAGGAAGAAAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 132

QY      184 ggcgc 187
        || |
Db      131 GGAG 128

RESULT      4
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
```

```

1  COUNTRY: USA
2  ZIP: 22313-0299
3  COMPUTER READABLE FORM:
4  MEDIUM TYPE: Floppy disk
5  COMPUTER: IBM PC compatible
6  OPERATING SYSTEM: PC-DOS/MS-DOS
7  SOFTWARE: Patentn Release #1.0, Version #1.25
8  CURRENT APPLICATION DATA:
9  APPLICATION NUMBER: US/08/232,463
10 FILING DATE:
11 CLASSIFICATION: 435
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: US/07/935,313
14 FILING DATE:
15 APPLICATION NUMBER: EP 91 114 300,6
16 FILING DATE: 26-AUG-1991
17 ATTORNEY/AGENT INFORMATION:
18 NAME: BENT, Stephen A.
19 REGISTRATION NUMBER: 29,768
20 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (703)836-9300
23 TELEFAX: (703)683-4109
24
25 INFORMATION FOR SEQ ID NO: 14:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 7218 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 IMMEDIATE SOURCE:
32 CLONE: pTZpc-F1s
33
34 US-08-232-463-14
35
36 Query Match 4.6%; Score 56.8; DB 1; Length 7218;
37 Best Local Similarity 2.4%; Pred. No. 0.00081;
38 Matches 7; Conservative 181; Mismatches 98; Indels 0; Gaps 0;
39
40 QY 63 ggggaaggaagcagagagatgcgagagcgcggaagcagagatcgcagagcgcgag 122
41 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
42 Db 1339 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1280
43
44 QY 123 ccgcgcgcgcagagagcgcagcgcagagagagcagagagatcgcagagccgcg 182
45 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
46 Db 1279 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1220
47
48 QY 183 cgcgcgcgcgcagcgcgtcgcagtcacgaatgacagggtcagaaaatcgcgagccaca 242
49 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 Db 1219 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1160
51
52 QY 243 gggacatcgtcagatcgcgcacaaatccgcgatctggtgagacgaaacgcgcaatg 302
53 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54 Db 1159 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1100
55
56 QY 303 aacgcacaaacctgatttcacagaatgagatccgcttcctgc 348
57 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 Db 1099 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1054
59
60 RESULT 6
61 US-08-770-379-20/c
62 Sequence 30, Application US/08770379
63 Patent No. 5849564
64
65 GENERAL INFORMATION:
66 APPLICANT: Chang, Yuan
67 APPLICANT: Bohenzky, Roy A.
68 APPLICANT: Russo, James J.
69 APPLICANT: Edelman, Isidore S.
70 APPLICANT: Moore, Patrick S.
71 TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
72 VIRUS
73 TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
74 NUMBER OF SEQUENCES: 20

```

```

? CLASSIFICATION: 530
? ATTORNEY/AGENT INFORMATION:
? NAME: Osman Ph.D., Richard A
? REGISTRATION NUMBER: 36,627
? REFERENCE/DOCKET NUMBER: UCB96-055
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415)343-4341
? TELEFAX: (415)343-4342
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2277 base pairs
? type: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
US-08-676-967-5

Query Match      .      4.5%; Score 56; DB 2; Length 2277;
Best Local Similarity 55.1%; Pred. No. 0.00095;
Matches 151; Conservative 0; Mismatches 120; Indels 3; Gaps 2

QY   64 ggaggaaggacgaaggatgcggagagcgcggaagagcagagactcgagagccgcgaagc 123
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   657 GAAGGAGGCCCGCCGGAGGAGGAGACATGAGAGGAGGAGAAGCAGCAGCAGCAGCA 716

QY   124 gcccgcgcgcgagggagcgagcgagcgagcgaggggagcagagagcgagtgcggagga-gccgcggy 182
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   717 CGAGCACGACGAGGAGGAGCGCGCTGTTCGACGACGACGACGACGAGGAGGAGAACATCGAGHG 776

QY   183 cggcgcgcgcgcagctcggttcagttccagaatgcaggggtccagaactcgg--cgagccac 240
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   777 CAAGGTGACCAAGCCCCGTCGACAGATCATCAAGAGCGCGCGTGAAGCGCCCCGCCCAA 836

QY   241 gagggacctggtggtgttccgcacaaatlccgcatcttgttgtagcagagactgtcaattg 300
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   837 GAGCAGCCACCAACACCAGGAGGAGGAGGACGCACCTTGAGAGGAGGACGACATCGAGCAGCGG 896

QY   301 ggacacgcccaaacctgatcttctacagaatatgag 334
    ||| | | | | | | | | | | | | | | | | | |
Db   897 CGAGGAGCTGGCCCGACGACGACGACACCGACCGCAG 930

RESULT      8
US-08-676-974-5
? Sequence 5, Application US/08676974
? Patent No. 5770422
? GENERAL INFORMATION:
? APPLICANT: COLLINS, KATHLEEN
? TITLE OF INVENTION: Human telomerase
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Science & Technology Law Group
? STREET: 268 Bush Street, Suite 3200
? CITY: San Francisco
? STATE: CA
? COUNTRY: USA
? ZIP: 94104
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/676,974
? FILING DATE:
? CLASSIFICATION: 530
? ATTORNEY/AGENT INFORMATION:
? NAME: Osman Ph.D., Richard A
? REGISTRATION NUMBER: 36,627
? REFERENCE/DOCKET NUMBER: UCB96-055
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415)343-4341
```



```
STATE: NY  
COUNTRY: USA  
ZIP: 11530
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT-US93-/06251
FILING DATE: 1993-06-30

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586

TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEX: 516-742-4366

INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 2301 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-78

Query Match 4.1%; Score 50.8; DB 6; Length 2301;
Best Local Similarity 62.7%; Pred. No. 0.012;

Matches 79; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

DY 64 ggaagaggacgagcggatgcggaggagcgaaacgagaccgcgagccgcgcaagc 123
| | | | |
DB 40 GGAGGACGAAGAAGAGCAGCAGAGGAGCGAGCACGAGGACAGAGGACAAGCAAAGGAGAGGA 99

OY 124 cgcgccgcgcgagggagcgagcgagggagcgagagcgagagcgagtccgagaagcgcccgcg 183
| | | | |
DB 100 GGAGGAGGAAAGAGGAGCAAGCAAGAACAGAGCCAGAGAGGAGSAGAGCAAGCAAGAGAGCA 159

OY 184 gccgcgcg 189
|| ||
DB 160 GGAGGG 165

RESULT 15
US-08-658-136-2/C
Sequence 2, Application US/08658136
Patent No. 607117

GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACIKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2000, 02:43:28 ; Search time 3331.48 Seconds
(without alignments)
2949.756 Million cell updates/sec

Title: US-09-431-843B-1
Perfect score: 2250
Sequence: 1 tgggcacgcacgcacccacg.....aaaaaaaaaaaaaaaaaaaa 2250

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pl3: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_sy: *
14: gb_un: *
15: em_fun: *
16: em_hum1: *
17: em_hum2: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_v1: *
30: gb_in1: *
31: gb_in2: *
32: gb_in3: *
33: gb_pl3: *
34: gb_pl4: *
35: gb_pr5: *
36: gb_htg1: *
37: gb_htg2: *
38: gb_htg3: *
39: gb_htg4: *
40: gb_htg5: *
41: gb_htg6: *
42: gb_htg7: *
43: gb_htg8: *

44: gb_htg9: *
45: gb_htg10: *
46: gb_htg11: *
47: gb_htg12: *
48: gb_htg13: *
49: gb_htg14: *
50: gb_htg15: *
51: gb_htg16: *
52: gb_htg17: *
53: em_ba1: *
54: em_ba2: *
55: em_htg1: *
56: em_htg2: *
57: em_htg3: *
58: em_htg4: *
59: em_htg5: *
60: em_htg6: *
61: em_htg7: *
62: em_hum3: *
63: em_hum4: *
64: em_hum5: *
65: em_hum6: *
66: gb_htg18: *
67: gb_htg19: *
68: gb_htg20: *
69: gb_htg21: *
70: gb_htg22: *
71: gb_htg23: *
72: gb_v11: *
73: gb_v12: *
74: gb_ba3: *
75: em_htg8: *
76: em_htg9: *
77: em_htg10: *
78: em_htg11: *
79: em_htg12: *
80: em_htg13: *
81: em_htg14: *
82: em_htg15: *
83: em_htg16: *
84: em_htg17: *
85: em_htg18: *
86: em_htg19: *
87: em_htg20: *
88: em_htg21: *
89: em_htg22: *
90: em_htg23: *
91: gb_pr6: *
92: gb_pr7: *
93: gb_sts1: *
94: gb_sts2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2208	98.1	2208	12 AF156878	AF156878 Rattus no
2	698	31.0	2363	11 AF172452	AF172452 Homo sapi
3	696.8	31.0	2423	11 AF172451	AF172451 Homo sapi
4	696.6	31.0	1676	11 AF172450	AF172450 Homo sapi
5	693.4	30.8	2483	11 AF172453	AF172453 Homo sapi
6	673.4	29.9	2423	11 AF109134	AF109134 Homo sapi
7	586.4	26.1	1232	11 AF172449	AF172449 Homo sapi
8	336.6	15.0	160241	35 HS88517	AL035669 Human DNA
9	329.4	14.6	6670	11 AF112980	AF112980 Homo sapi
10	86.6	3.8	2830	30 AF027972	AF027972 Nephtia c
11	81.6	3.6	14999	31 AF218621	AF218621 Nephtia c
12	75.8	3.4	1278	32 PFACSAAD	J02751 Plasmodium

Db 661 CACCTTGAGGACCGGGGACCGGGTCTGTATGCCGTGCACAGAACTTCCAGCCGCGCTTC 720
Oy 721 caaatctgaacagccacagccacaaactgctgattacacgactctcaactgactg 780
Db 721 CACATCTCAACAGCCACAGCCCAACAACTGCGATTACAGCATCTCTCAAGTCACTG 780
Oy 781 ggtgagctgggtgtaagaactacacagccctggtcgcgtctctctggaggagacc 840
Db 781 GGTGAGCTGGGCTTAGAACACTACAGGACCCCGTGTCTTCTTCTGAGAGGAGAC 840
Oy 841 ctgtgacagcaaaactccagcgtgagccagatgaccttgactactctgttgcgt 900
Db 841 CTGTGACAGCAAACTGCGCCAGCGGTGCGCCAGAGTGGCCCTGACACTCTCTGTCTGCT 900
Oy 901 ggtgctcccgccacacagccggtgagctgtgtacttctgctggagaaactcaagcct 960
Db 901 GGTGCTCCGCGCCACAGGCGCGGGAGCTGTGTACTTGTCTGAGGAGCACTTCAAGCT 960
Oy 961 cgcagagattgtctggtgggccccgtgacaagctgacagattcaagccacagacata 1020
Db 961 CGCCAGAGTTGTCTGGGGGGCCCGTGACAGCTGCGGAGTTCAAGCCCAAGACATA 1020
Oy 1021 cccagacagctgacagggagacagggagcagaataagatgagggctccagggaccctcc 1080
Db 1021 CCCAGCCACTGACGGGACCAAGGGACGCGAGTAAAGATGAGGGCTCCAGGACCCCTCC 1080
Oy 1081 caagaggtctgagccacaggggtgagacctgtgactctgagagagacctggtgtggagacgt 1140
Db 1081 CAAAGAGGCTGGGACCCAGGGGTGGAGCTGTGATCTGGAAGGAGCTGTGTTGGGCACT 1140
Oy 1141 ggaacagctgagatccctcactgctgtaacacaaagccctcagatgagggagaccttgcgt 1200
Db 1141 GGAACAGCTGAGAGATCCCTCACTGCTGAACACAAAGCCCTCAGATGAGGGGAACCTTGGAT 1200
Oy 1201 gggagacagaggtatgaaactgaactgacctgagatcccaagagagacaaagaagtgtg 1260
Db 1201 GGGAGCCAGAGGGATGAAAGCTAAAGTCCCTGAGTCCCAAGAGAGCAAGAAAGAAAGTTGG 1260
Oy 1261 gagggagacagagcagagagagatcccaagggagagagatcccaaggggtgtctcctaagta 1320
Db 1261 GAGGGGAACAGGCAAGAGAGAGTCCCAAGGGAGGAGATCCCAAGGGGTGTCTCTGAGGTA 1320
Oy 1321 gagaacatgccccttaaccttgagaggtgtgccccttaagccctatcagccagagacagcag 1380
Db 1321 GAGAAATTTGCCCTTAACCTTGAGAGAGTGTGCCCTTAGGCCATACGACGAGAGCCAGG 1380
Oy 1381 gaggtctgacagccctgtctcgtgtgagcagaggtgtgctaataggttaagaagcagaggaag 1440
Db 1381 GAGGCTGAACCGCCCTGTCTGTGTGCGCCAGGGGTGGCTAATGAGTAAAGAAAGCGGAGGAAG 1440
Oy 1441 gtagagagagagaggtgtagaggtgtagtagtaactaactcaactcaactgagccagctgtgc 1500
Db 1441 GTGGAGGAAGGGGCTGAGAGGTATGAGAGTACAGTAACTCAAAATGAGAGGCCAGTGC 1500
Oy 1501 ctgacctctacccctctcagaggtgtcctcagagcccaaaagatgaggaatgagccagagac 1560
Db 1501 CTGCTCTTACCCCTTACAGAGTGTCTGAGGCCCAAAAGAGTGGGAGTGGGACAGAGAC 1560
Oy 1561 tcaaacagccagaggttggggcagagagattccaaagccaggtgggcccagaggaattccaaac 1620
Db 1561 TCAAAACAGCCAGAGTGGGGCAGAGATTCCAAAAGCCAGGTGGGGCCGAGGATCCAAAC 1620
Oy 1621 agccaggtggggctgtagagagcccaaacagcagctgagccagagagaccccaaacagcag 1680
Db 1621 AGCCAGGTGGGCTGAGAGACCCAAACAGCCAGGTGGGGCCAGAGAGCCCAAAACAGCCAG 1680
Oy 1681 gtcgagagagagagacacacagccaggttcgagccagagagaccccaaacagcaggttcgag 1740
Db 1681 GTGCGGACAGAGAGCCAAACAGCCAGGTGGGGCCAGAGAGCCAAACAGCCAGGTGGG 1740
Oy 1741 ccagagagacacacacagccaggtgtgtggggcagagagcaagctgtcctctaagaagccctgtg 1800
Db 1741 CCAGAGGAGCCCAAAACAGCCAGGTGTGTGGGGCCAGAGCAAGGTGCTCTAAGAGCCCTGTG 1800

Oy 1801 gagagccctgactctgacactatgggaacctccactgagatgagtcagagaggtgcaag 1860
Db 1801 GAGGACCTGACTCTGACACTATGGGAACCTCAGTGAATGAGTCAGAGAGTGGCAAG 1860
Oy 1861 attgagcctctgctgaaccccccaagccttagaggtgacatctcagctactcaagcca 1920
Db 1861 ATTGAGGCTCTGCTGGAACCCCAAGCCTTAGAGCTGACATCTGCTCTACTGAGCCCA 1920
Oy 1921 ctgagggggttctcagtcagagctctgcgttagagctcttctgtgtgccccagatgc 1980
Db 1921 CTGAGGGGGTTCGTGATCCAGAGCTGTGCGGTGAGCTCTTGTGTGGTCCCAAGTGC 1980
Oy 1981 tggcctccctgaaggtgacactgaggtgagccagagagagagagagagagagagagag 2040
Db 1981 TGGCTCTCCCTAGTGTGACTGAGTGGCCACAGAGGAGCTGAGGGCCCTGCGCTCAGG 2040
Oy 2041 gaaagccaaagcctcacaacccctcactcaactgctgctcctccacagccctcga 2100
Db 2041 GAAAGCCAAAGGCTTCAGAACCCCTTCACTCACTGTGTCTCTCCACTGCGCTCTGA 2100
Oy 2101 gccctgctgtgacagaccctcagaggtctagagagagagagagagagagagagagag 2160
Db 2101 GCCCTGCTGTGATCAACCCCTAAGGGCTAGAGGAGAGGGGCTCTTCAATAGTCTG 2160
Oy 2161 gccagtgagagccttcttgataaactcttagagcttgcacaaaaa 2208
Db 2161 GCCAGTGAAGGCTTTTGTGAATAAATCTTGTAGCTTTGTCAAAAAA 2208

RESULT 2

AF172452

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .2363

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q13.3"

/tissue_type="placenta"

34. .2007

/note="[Met-5]-enkephalin binding protein; OGRF;

alternatively spliced"

/codon_start=1

/product="opioid growth factor receptor"

/protein_id="AA64405.1"

/db_xref="GI:7595305"

/translation="MDPPDCDSMEEDPDAEDPCDGEAARADADDEDE

SEBRARPSFSQSMRTGSRNKRATRMCKRYRHPDVLVRDCNGDTPNISTFNRER

FLNPGCFIEDILQMDNDYDLDEDNHSYIQWLPPLRPGVNMWAKPLTLEVEVFKS

QEI0ERLVRAVEYELMGFGYGLRDRGTVGRQANTOKRPONLNMWSHNNLRITRLIK

SPECILSEHQAPLVRFLEETLVRRLPGVRSALDQYFMFAVRCROROLVHFAWE

HFRPRCKFVWGPDQKLRRFPSSLPRLDESRVVEEGSGDPDDEASITGCRICGPER

TITLE Direct Submission
JOURNAL Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA

FEATURES
source
Location/Qualifiers
1..1676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue_type="Placenta"
34..1419
/note="[Met-5]-enkephalin binding protein: OGF:
alternatively spliced"
/product="opioid growth factor receptor"
/protein_id="AA66403.1"
/db_xref="GI:759301"
/translation="MDPDCSTWEDEDADEDECEDEAGADADADEDEE
SEEPARPSFOSRMTRRMVATRDRCRHRNPDLYVERDCNDPTNLSFVREIR
FLPNCFTEDIONMTDNYDLLENHAYLOMTFPRREGVNHAKPULREVEYKSS
OEIOIRLVRAYFLMAGFYGRIEDNGTGTGVAQNTOKRFOILNKRSHNNTITILK
SPELSLHFQAPLVRFLEETLVRRLEPVQRSLDYFMFAVRGRQRLVHFAWE
HFRPKKFWMGQDLRRFKRSPSLPDEGSKYBESPGDPDHEASTOGRTGPEH
SKGGRVDEGPQPSRVEYQDAGPLERSGDEGDEPRPELSPKSKRKLEISR
EOPTEPQPSAVERKIALNLEGCALSQSLRTGTQVGGQDPPEASCPCCRWGLR
SCCGLPSSAS"

CDS

BASE COUNT 340 a 501 c 575 g 260 t
ORIGIN
Query Match 31.0%; Score 696.6; DB 11; Length 1676;
Best Local Similarity 72.3%; Pred. No. 3e-145;
Matches 944; Conservative 0; Mismatches 334; Indels 27; Gaps 2;

QY 135 ctcctgctgtcagatgacgaacggaatcgcattccaccctggaggagagagcga 194
DB 48 CTGGCACTCCACTGGAGAGGAGGAGATGCGAGCAGCGGAGAGAGAGACTG 107
QY 195 ggaagatgacgaagatgacgaagcgaatgacacgagtgagacagggagagatga 254
DB 108 CAGAGACGCGAGCGCCGCGCGAGGAGCGGAGCGAGGAGGAGAGAGATGC 167
QY 255 cggcagcggagagagcagcggcgaagcctgttcacagtcgaagatgacagagaa 314
DB 168 GGAGGAGCG 227
QY 315 ctggcgctgtatgacgaagatgacgaagcgaagcgaagcgaagcgaagcga 374
DB 228 CTGGCGAGCGACGAGGAGATGTGTAGGTGCGACAACTATCCGAGATCTGTGAAGC 287
QY 375 agacgtcaatggtgacatgacacgtgacgttcacaaatgagatgacgttcacac 434
DB 268 AGACTGCAATGGGAGACGCCCAACCTGATTCTACAGAAATGAGATCCGCTTCTGCC 347
QY 435 aaatggggcctcctcagagacatcttcacgaactggaagaaacacatgacacctgga 494
DB 348 CAACGGCTGTTTCAATTGAGACATCTTCAGAACTGAGCGGACACTATGACCTCTTGA 407
QY 495 agagaataactctcattcattcagtgctgtttcctctcgaggagacgaagatgagca 554
DB 408 GGACATATCTCTACATCCCACTGCTGCTTCTCTGGAGAACGAGGAGATGCGCA 467
QY 555 cgcacgaacccctcaccctggaagaggtgagacatttaaaagctccaagagatgagga 614
DB 468 TGCCAAAGCCCTCAAGCTCAAGGAGAGTGAGGTGTTTAAAGCTCCCAAGGAGATCCAGA 527
QY 615 ggcctgttcggggcctatgagctcattgctgttctcattggttccacctgagagcag 674
DB 528 GCGGCTTGTCCGGGCGCTAGAGCTCATGCTGGGCTTCTAACGGGAGCCGCGGAGGAGCAG 587
QY 675 gggacgggtgtgtatgctgtgacaaatctcagcgcggcttcacgaactggaacag 734

DB 588 AGGCACGGGCGACGTTGGGCCGACACAGACTACCAAGAGCGCTTCCAGAACTGAACTG 647
QY 735 ccacagccacacaaacactgctlatatcacgcatctcgaactcgaactggtgagctggt 794
DB 648 GCGGACGACCAACACCTCCGATCAGACAGCATCTCCTCAAGTGGCGTGTGAGTGAGCT 707
QY 795 agaatactacccaagacccctgtctcgttcttctctgagagagacacctgttacaacaa 854
DB 708 CGAGCACTTCCAGGCGCCACTGTGCTCGCTTCTCTGAGAGAGAGCGTGGCGCGGGA 767
QY 855 actgcccagctgtgcagagatgcccctgactactctctgtctgtctgtcgtcgcga 914
DB 768 GCTGCGGGGGGTGCGCGCAGAGTCCCTTGACTTATGTTGCTCCGCTGCGCTCCGACA 827
QY 915 ccagcgcggagagctgtgtacttctgtcctgtggagacacttcaagcctcgccgagattgt 974
DB 828 CCAGGCGCGCCAGCTGTGGCTGCTGCTGCGAGACACTTCGCGCCCGCTCAAGTTGCT 887
QY 975 ctgggggccccgtgacaaagtgcgagagattcaagccccacatccccagccactgac 1034
DB 888 CTGGGGCCCCAAGCAAGCTGCGAGGTTCAAGCCCACTCTGTGCGCATCCGCTCGA 947
QY 1035 gggaccagggcagcagataaagatgaggtcgaaggagacccctcccaagagctgtgcac 1094
DB 948 GGGCTCCAGGAAGGTGAG 1007
QY 1095 ccaaggtctggaactgtgtatctgtaaggagacgtgagtgggagacagtgaacagctgaga 1154
DB 1008 CCAGGCTCGGACCTGTGGCGCAGAGCATAGCAAGGCTGGGGGAGG---TGACGAGAGG 1064
QY 1155 tccctcactgtgacacacaaagccctcagatgggggagaccttggatgagagagagga 1214
DB 1065 GCCCAGCGACGAGAGCGTGGAGCGCCAGATGCGGAGCCCTGAGAGAGAGAGAGAGAG 1124
QY 1215 tgaag-----ctgaagtcctctgagtcaccaagagagacagaa 1250
DB 1125 TGAGGAGAGGGGCGACGAGGAGATAGCGCGAGACCTTTAAGCCCCAAGAGAGCAAGAA 1184
QY 1251 aaggaattgagagagagacagagcagagcaggtgtccaggggagagcagatcccaaggtgt 1310
DB 1185 GAGGAAGCTGAGCTGAGCGCGCGGAGCAGCCGCCCAAGAGAGAGAGAGAGAGATGC 1244
QY 1311 ctcgtgagtagaagaattgccccttaaccttgaagagtggtgcccctgaccttaagcca 1370
DB 1245 CTCAGAGGTGAGAGAGATCGCTGTGAATTGTGAGAGGTGTGCCCTCAGCGCAGGCACT 1304
QY 1371 ggaagccagggaggtcgaacgcctgtctgtgtggtgccaaggtgtgc 1415
DB 1305 CAGGACGGGGGACCCAGGAGAGTGGCGGTCAAGACCTTGGGGAGGC 1349

RESULT 5
AF172453 2483 bp mRNA PRI 19-APR-2000
LOCUS AF172453
DEFINITION Homo sapiens opioid growth factor receptor mRNA, complete cds.
ACCESSION AF172453.1 GI:7595306
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2483)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Cloning, sequencing, chromosomal location, and function of cDNAs
encoding an opioid growth factor receptor (OGFR) in humans
Brain Res. 856 (1-2), 75-83 (2000)
JOURNAL MEDLINE
2 (bases 1 to 2483)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Direct Submission
TITLE Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University

[illegible]

Db	623	GCCTTCCAGAAACCTGAACATG--CGCAGCCCAACAACACTCCGATCACACAGCATCCTCA	681
QY	774	gtatcgtggtggtgacgttggtctagaacaactacacaggaaccccttgctcgtctctctctgga	833
Db	682	GTGCTGTGGGTGACCTGGGGCTCTCGAGCACTTCCAGGGCGCGCTGTGCTCTTCTTCTGGA	741
QY	834	ggaagaccttgtacacagacaaactgtccacgctgtgcgcacagatgtccctgtgacctctct	893
Db	742	GGAGACCTGTGTGCGCGGAGCTGCCGGGGGGTGGCGGAGTGGCCCTGGACTACTTCAT	801
QY	894	gttcgtgtgtgctgtgcgcgcacacaggtccgggtggtgtgtgtacttgcctgtggagactt	933
Db	802	GTTCGCGGTGTGGTGTGCGACACCCAGGCGCGCCGCTGTGTGTGCACTTGCCCTGGAGCACTT	861
QY	954	caagcctctgcgcgaagtgttctctgtgtggccccgtgaacaagtgcggagattcaagcccca	1013
Db	862	CCGGCCCCCGCTGCAATTTGTTGTGTGGGGGGCCCCAAGACAACTGTGGAGTTCAAGCCCA	921
QY	1014	gacacataccccagccaactgtacggtgacaggtgcgaagataaagatgtatgaggtctcagga	1073
Db	922	CTCTCTGCGCGCACTCCCTCTGAGGGCTCCAGGAAGTGTGGAGGAGGAAGAACCCCGGGGA	981
QY	1074	ccccctcccaagaggtctgtgcacccaggtgtggaacctgtgatactgtgaaggtacctgtgtg	1133
Db	982	CCCCGACCAAGAGGCCACGACACCCAGGGGTGGACCTTGTGGCCAGAGCAATAGCAAGGTGG	1041
QY	1134	ggaacagtgaacaagctgtgaagatccctactgtctgaaacaaagacccctcagatgtgtggaac	1193
Db	1042	GGGCGAAGG---TGGAGGAAGGGGCCCCAGCCAGAGAGCTGTGGAGCCCAAGATGTGGGACC	1098
QY	1194	cttggatgttggaaccacgaaggtatgaag-----ctaatgtccct	1229
Db	1098	CCTGGAAGAGGACCAAGGGGATGTAGGCAAGGGGCCACGGGGAAGATAGGCCCGAGCCCTT	1158
QY	1230	gagttcccaagtgagagcaagaagaaaggaagtttgtaggtggaacaggtcagagacaggtccag	1289
Db	1159	AAGCCCCAAGAGAGACAAAGAAAGAGAGCACTGTGAGCTGAGCCGCGGGAGACACCGCCAC	1218
QY	1290	ggaagcagaatccccacgggtgtctctgtgaggttaagaaatgtcccttaacctgtgaagatgt	1349
Db	1219	AAGGCGAGGCGCTCTGAGAGTGTCTCAGAGGTGGAGAAGATCGCTTGAAATTGTGAGAGGTG	1278
QY	1350	tgaccttaagccctatcagc-----caggagcccg	1379
Db	1279	TCCCTTCAGACGAGGGGAGGCTCTGAGAGCGGGAGCCACAGAAAGTGGGGCGGTCAAGACCTCTG	1338
QY	1380	ggaagctgtgaacggccctgt-----tcctgttgccaggtgtgtctaatagtgaagaa	1430
Db	1339	GGAGGCAAGTGCAAGGCTCTGCCCAACCCCTGGGAACCAAGGGTGGCCGACAAAGTAGAGGA	1398
QY	1431	gagggaggaagtgtagggaggaaggtctgtgaaggtatgtgaagtgt-----cagtaacactca	1484
Db	1399	GGGGAAGAAAGTGTGAGTGAAGGTGTGTGGGGACACACTCTCTCGGTGGCCAGTGTGTGTGCCCA	1458
QY	1485	aatgcaggtccaggtgcctctgcctctcaacctcaagatgttccgtgaagcccaaaagatgtg	1544
Db	1459	GACCTTTGGCGCTTGGCGGGGTCCCGCCGACCAACGGGGGACCCCAAGGCTGTGACACAGTGA	1518
QY	1545	gaatgtgcacagga	1559
Db	1519	GAAAGGGGTGAGGA	1533
RESULT	7		
AF172449	AF172449	1232 bp	mRNA
LOCUS	DEFINITION	Homo sapiens clone 127	optoid growth factor receptor mRNA, complete cds.
ACCESSION	AF172449		
VERSION	AF172449.1	GI:7595298	
KEYWORDS	human.		

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
RP5-885L7 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-885L7 The true left end of clone RP4-563E14 is at 140192 in this sequence. The true right end of clone RP11-93B14 is at 44512 in this sequence.

FEATURES

```
source
1. 160241
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="20"
   /map="q13.2-13.33"
   /clone="RP5-885L7"
   /clone_1lb="RPCI-5"
2. 64
   /note="L1MA9 repeat: matches 6210. .6254 of consensus"
repeat_region
65. .333
   /note="AluIo repeat: matches 1. .287 of consensus"
repeat_region
354. .397
   /note="L1MA9 repeat: matches 6254. .6302 of consensus"
4198. .5157
   /note="12 copies 80 mer 61% conserved"
   /complement(4225. .5153)
   /note="Tandem repeat. region contains two forced joins and single clone. Assembly consistent with restriction digest."
repeat_region
4229. .5114
   /note="443 copies 2 mer cc 59% conserved"
repeat_region
4231. .5166
   /note="234 copies 4 mer tccc 60% conserved"
repeat_region
4240. .5151
   /note="19 copies 48 mer 61% conserved"
repeat_region
4248. .5159
   /note="38 copies 24 mer 61% conserved"
4264. .5163
   /note="15 copies 60 mer 61% conserved"
repeat_region
4286. .5152
   /note="17 copies 51 mer 59% conserved"
4308. .5171
   /note="27 copies 32 mer 61% conserved"
repeat_region
4317. .5156
   /note="21 copies 40 mer 62% conserved"
repeat_region
4344. .5142
   /note="17 copies 47 mer 59% conserved"
repeat_region
4348. .5145
   /note="42 copies 19 mer 60% conserved"
repeat_region
4431. .5144
   /note="21 copies 34 mer 60% conserved"
4456. .5148
   /note="21 copies 33 mer 60% conserved"
repeat_region
4486. .5151
   /note="9 copies 74 mer 69% conserved"
4530. .5087
   /note="18 copies 31 mer 71% conserved"
repeat_region
4533. .4822
   /note="5 copies 58 mer 77% conserved"
repeat_region
4720. .5137
   /note="11 copies 38 mer 61% conserved"
repeat_region
4833. .5064
   /note="4 copies 58 mer 79% conserved"
repeat_region
5051. .5164
   /note="19 copies 6 mer tctctc 63% conserved"
repeat_region
5934. .6864
   /note="L1M4 repeat: matches 4509. .5488 of consensus"
6974. .7049
   /note="4 copies 19 mer 80% conserved"
repeat_region
7224. .7451
```

```
repeat_region
7892. .8074
   /note="6 copies 38 mer 98% conserved"
repeat_region
8674. .8966
   /note="MER68A repeat: matches 68. .210 of consensus"
repeat_region
10876. .10962
   /note="AluY repeat: matches 5. .296 of consensus"
repeat_region
11974. .12025
   /note="MIR repeat: matches 86. .162 of consensus"
repeat_region
14394. .14478
   /note="13 copies 4 mer ctca 73% conserved"
repeat_region
15824. .15891
   /note="5 copies 17 mer 77% conserved"
repeat_region
16266. .16577
   /note="17 copies 4 mer tccc 67% conserved"
repeat_region
16578. .16866
   /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region
16581. .16844
   /note="17 copies 17 mer 65% conserved"
repeat_region
16582. .16891
   /note="11 copies 24 mer 68% conserved"
repeat_region
16583. .16879
   /note="10 copies 31 mer 66% conserved"
repeat_region
16585. .16840
   /note="9 copies 33 mer 66% conserved"
repeat_region
16586. .16875
   /note="16 copies 16 mer 66% conserved"
repeat_region
16590. .16874
   /note="5 copies 58 mer 72% conserved"
repeat_region
16593. .16872
   /note="15 copies 19 mer 64% conserved"
repeat_region
16598. .16837
   /note="7 copies 40 mer 67% conserved"
repeat_region
16603. .16842
   /note="5 copies 48 mer 73% conserved"
repeat_region
16638. .16901
   /note="3 copies 80 mer 77% conserved"
repeat_region
16664. .16885
   /note="44 copies 6 mer aaaga 62% conserved"
repeat_region
16670. .16904
   /note="3 copies 74 mer 78% conserved"
repeat_region
16848. .16899
   /note="5 copies 47 mer 77% conserved"
repeat_region
17287. .17411
   /note="13 copies 4 mer gaag 100% conserved"
repeat_region
17447. .17486
   /note="MERSA repeat: matches 35. .165 of consensus"
repeat_region
18303. .18502
   /note="10 copies 4 mer ccat 92% conserved"
repeat_region
18307. .18498
   /note="5 copies 40 mer 66% conserved"
repeat_region
18317. .18456
   /note="4 copies 48 mer 67% conserved"
repeat_region
18319. .18486
   /note="3 copies 60 mer 74% conserved"
repeat_region
18320. .18495
   /note="7 copies 24 mer 66% conserved"
repeat_region
18321. .18496
   /note="44 copies 4 mer aac 64% conserved"
repeat_region
18325. .18486
   /note="11 copies 16 mer 64% conserved"
repeat_region
18327. .18486
   /note="9 copies 18 mer 66% conserved"
repeat_region
18497. .18546
   /note="8 copies 20 mer 66% conserved"
repeat_region
19329. .19433
   /note="25 copies 2 mer ac 77% conserved"
repeat_region
19455. .19590
   /note="LTRL6C repeat: matches 286. .387 of consensus"
repeat_region
19457. .19604
   /note="4 copies 34 mer 88% conserved"
repeat_region
19546. .19625
   /note="2 copies 74 mer 81% conserved"
repeat_region
   /note="4 copies 20 mer 76% conserved"
```

repeat_region	19839..19954	/note="2 copies 58 mer 87% conserved"
repeat_region	20020..20103	/note="MR repeat: matches 51..137 of consensus"
repeat_region	20704..20805	/note="3 copies 34 mer 90% conserved"
repeat_region	20710..20811	/note="2 copies 51 mer 90% conserved"
repeat_region	20722..20806	/note="5 copies 17 mer 89% conserved"
repeat_region	21349..21410	/note="2 copies 31 mer 93% conserved"
repeat_region	21584..21654	/note="LR16A repeat: matches 88..158 of consensus"
repeat_region	21779..21852	/note="Alu repeat: matches 238..311 of consensus"
repeat_region	24787..24847	/note="MR91B repeat: matches 1..63 of consensus"
repeat_region	24855..25064	/note="MR10 repeat: matches 107..348 of consensus"
repeat_region	25853..26065	

Query Match	15.0%;	Score 336.6;	DB 35;	Length 160241;
Best Local Similarity	66.4%;	Pred. No. 9e-65;		
Matches 596;	Conservative	0;	Mismatches 229;	Indels 72; Gaps 5

QY	732	caagcaagagcccaaaacaaactgcgfatataacgcagctctcaagatcaatcaggtgtgagcttgg	791
Db	61404	caagcgacagcccaaaacaaactctccgatatacaacgcattcctaagctgcgtggatgaactggg	61463
QY	792	ctlaagaactactaccagagcaccctctgtccgctctctctctctgaaagagaaacctgtgtacaaca	851
Db	61464	ccttcagacactttccagagcgccctgctgctccgcttcttctgaaagagaaacctgtgtacaaca	61523
QY	852	caaaactgtcccagcgctgtgagccagagatgagccctggactacttctgttgcgtgtgctgtccg	911
Db	61554	ggagctgtccgggggggtgcggcagagatggcccttgactacttccatgttcaggttcgcccgtgcg	61583
QY	912	gcaaccagcgcccgagagctctgtgtactcttgcctgagagcaacttcaagctctgcgcgaagatt	971
Db	61584	aaacagagcgccgcacgctggtgcaacttgcgcgtggagagcaacttccgcccgcgtgcacattt	61643
QY	972	tgtctgtggtgtcccgctgacaaagctgtgcgaatataagcccccagaccataccccagcaact	1031
Db	61644	cgctcgtggggggcccccagaaagcttccggagatttcamaagccccaagcttctcgtcccaactccct	61703
QY	1032	gacggagaccaggtgcagagcagataaagatgtgaggtgtccagggaaaccttcacaaaggtctg	1091
Db	61704	cgaggggcttccaaagaaaggtggagagaaagaaacccccggggaaacctccagacaaagagccag	61763
QY	1092	caaccagaggtctcgagactgtgtgatcttgaagaggaacctgtgtgtggtgagacagltgaaacgctga	1151
Db	61764	caaccagaggtctcgagactgtgtgatcttgaagaggaacctgtgtgtggtgagacagltgaaacgctga	61820
QY	1152	ggatctccctcactgtctgaaacacaagaacctctcaagtatgtgtgggaacctgtgtatgtgaaacag	1211
Db	61821	ggggccccaagccacgcgaacgctgacaccccaagatccggagaccctctgtgagagagacccaagg	61880
QY	1212	ggatgtgaag-----ctaaagtccctgaagtcaccaagagaaagcaaa	1247
Db	61881	ggatcagagccagggggccacaggggaagatgaagccggagacccttaaacccccaagaagaagcaaa	61940
QY	1248	gaaaaagaaagtctgtgaggtggaacaggtcaggtgagctcccaaggggtgaggtcagaatccccaggg	1307
Db	61941	gaagaggaagctgtgaggtggaacaggtcaggtgagctcccaaggggtgaggtcagaatccccaggg	62000
QY	1308	tgctctcaggtatagaagaatattgccttcaaccttgaagagatgtgccttcagccatcatcag	1367
Db	62001	tgccctcagaggtgcgaagaagatgcctctgaattttggaggggtgtgcccttcaagccaaagcgacg	62060
QY	1368	c-----caggagcccaaggtgagcttgaacgcgctgt	1397
Db	62061	ccttcagagccgggacccacagaagtggcggcttataggaaccttgaggagggagagtcacaccccttg	62120

Oy	1398	tcct-----gtggccagaggtgtgcaataaggttaagaaagcgagaggaagttgaagga	1448
Db	62121	CCGCGACACCCCTGGAGCCAGGGTGGCGACAGAGTGAAGAAAGCGAGAAAGTGGATGA	62180
Oy	1449	agggagctcgaagtgatgtagtgatgt-----cagtaacactcaatgcagcagtcacct	1502
Db	62181	GGGGTCGTGGGGACAGAGTCGTGGGTGGCGACAGTGTGTGTCGCCAGACCTTTGGCGGG	62240
Oy	1503	gctcctaccctcttcagagtgctcttcagagcccaaaagatgtagatgtagcccaagaga	1559
Db	62241	GTCCCCCTGCCCCATCGGGGACCCCCAAGGCTGTGAGACAGTGAAGAAAGCGGGTTGGGA	62297

() ()

RESULT		9			
AF112980	AF112980				
LOCUS	AF112980	6670 bp	DNA	PRI	10-JAN-1999
DEFINITION	Homo sapiens	7-60 gene,	complete cds.		
ACCESSION	AF112980				
VERSION	AF112980.1	GI:4139271			
KEYWORDS					
SOURCE	.human.				

ORGANISM Homo sapiens
Homio sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE	1 (bases 1 to 6670)
AUTHORS	Takanosu,M., Liu,J. and Mayne,R.
TITLE	Genomic Structure of Human Gene 7-60
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 6670)
AUTHORS	Takanosu,M., Liu,J. and Mayne,R.

JOURNAL Submitted (10-DEC-1998) Cell Biology, University of Alabama at Birmingham, VH 302, Room 605, Birmingham, Alabama 35294, USA

FEATURES Location/Qualifiers

```

SOURCE
1. .0010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
join(<206. .968.906. .984.2235. .2313.3147. .3213.4125. .4274
mRNA

```

join(206, .268, 906, .984, 2235, .2313, 3147, .3213, 4125, .4274, 4891, .6309)	
--	--

/translation="MTGSRNMRRATRDMCRYPHNPDLVERDCNGDPTNLSFYRNEIIF
 /ocid="1.137272"
 LPNCGFEJLITQANTDNDYDLEEDNHSYIOWLFTPLREGVYMWKAPLPLREVEYFSQ
 EIOGRIVRAYAAHAGLLRDLPGCFRHGSGPSTLPALPEPELRSNNIRILRSQ
 LGELIGLEFOPALVRFLEESIYLRRELPGROSADLFYMFVAGELRHRODLOVYHAWH
 FRPCKCFVWGPQDKLRRFKRSSLPHPLEGSRAVEESPPDPEHASTQRTCTGCPHS
 KGGGRVDEGPQPSRVSEFPODAPPELRSOSDEGGEDBPPEPLSKESKKRLELSRE
 OQPPPEPQASAEYERIALNLBEGCALQGSILRTGTOVGQDQDEEAVOPORAGAV
 ADKRRRKRYDEGASDAVAAGQIALLAGSAVSGHKAQCSNMGVEDIEGRTG
 PKSETPGSPSETPEGPSPAGPADDEPASESTPEGPSPAGPADDEPASESTPEGPRA
 GPADDEPAKTPSETPGSPAGPTDDEPASESTPEGPSPAGPADDEPASESTPEGPRA
 PAGVAGGEPAESPSETPEGPSPAGCTDDEPAPKAGAAELQDAVESSAKSGNP"
 BASE COUNT 1367 a 1998 c 2169 g 1136 t
 ORIGIN

Query Match	14.6%	Score 329.4	DB 11	Length 6670
Best Local Similarity	66.1%	Pred. No. 3.3e-63		
Matches 590; Conservative	0	Mismatches 231	Indels 72	Gaps 5

[illegible]

```

Db 4951 GAGACACTCCAGCGCGCGCTGCTCCCTCTGAGAGAGAGCCTGGTGGCGCGGAG 5010
Qy 856 cgtcccaaggttgcagagagtgccctggactacttccttgcgtgcgcgcac 915
Db 5011 cttccctgggggtgggagagtgccctggacttatttattgtccctcgtccgctccacac 5070
Qy 916 cagcgcggagagcttctgaacttctgctggagacactcaagcctgcgcgagattgtc 975
Db 5071 CAGCGCGCGCAAGTGTGCACTTCGCTGGAGACATTCGGCGCCGCTGCAAGTTCTGTC 5130
Qy 976 tggggcccccgtgacaagcttgcggagattcaagccccaacatacccaagcactgaac 1035
Db 5131 TGGGGCGCCCAAGACAGCTGCGAGGTTCAAGCCAGCTCTGCTGCGCATCGCTCGAG 5190
Qy 1036 ggaaccagggcagcagatataagatgaggtctcagagagacccctcccaagagctgcacc 1095
Db 5191 GGTCCAGAGAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5250
Qy 1096 cagagctgcagacttgcagctgcagagagagagagagagagagagagagagagagagag 1155
Db 5251 CAGGCTCGGAGACCTGTGGCGCAGAGCAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5307
Qy 1156 cccctcactgtcaacaagaagccctcagatgagggagagagagagagagagagagagag 1215
Db 5308 CCCCAAGCCAGGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5367
Qy 1216 gaag-----gtaagtccttcagtcagagagagagagagagagagagagagagag 1251
Db 5368 GAGGCGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5427
Qy 1252 aggaagattgagagagagagagagagagagagagagagagagagagagagagagagagag 1311
Db 5428 AGGAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5487
Qy 1312 tctgaggttgcagagagagagagagagagagagagagagagagagagagagagagagag 1368
Db 5488 TCAGAGGTGGAGAGATCCCTCTGAATTGGAGGGGTGTGCCCTCAGAGAGAGAGAGAG 5547
Qy 1369 -----cagagagagagagagagagagagagagagagagagagagagagagagag 1401
Db 5548 AGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5607
Qy 1402 -----gtggcagagagagagagagagagagagagagagagagagagagagagagag 1452
Db 5608 CAACCCCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5667
Qy 1453 gctgagaggttgcagagagagagagagagagagagagagagagagagagagagagagag 1506
Db 5668 GCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5727
Qy 1507 cctaccccttcagagagagagagagagagagagagagagagagagagagagagagagag 1559
Db 5728 CCGTCCCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5780

```

```

RESULT 10
AF027972 2830 bp mRNA INV 05-FEB-1998
LOCUS AF027972
DEFINITION Nephila clavipes flagelliform silk protein (Flag) mRNA, partial
cDS.
ACCESSION AF027972
VERSION AF027972.1 GI:2833646
KEYWORDS
SOURCE
Nephila clavipes.
Nephila clavipes.
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
REFERENCE
1 (bases 1 to 2830)
AUTHORS Hayashi,C.Y. and Lewis,R.V.
TITLE Evidence from flagelliform silk cDNA for the structural basis of
elasticity and modular nature of spider silks
JOURNAL J. Mol. Biol. 275 (5), 773-784 (1998)

```

```

MEDLINE 98153262
REFERENCE 2 (bases 1 to 2830)
AUTHORS Hayashi,C.Y. and Lewis,R.V.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1997) Department of Molecular Biology, University
of Wyoming, P.O. Box 3944, Laramie, WY 82071-3944, USA
FEATURES
source
1..2830
/organism="Nephila clavipes"
/db_xref="taxon:6915"
1..>2830
/gene="Flag"
219...>2830
/gene="Flag"
/codon_start=1
/product="flagelliform silk protein"
/protein_id="AAC38846.1"
/db_xref="GI:2833647"
/translation="MGKGRHDTKAKAMQVALASSIAELVTAESSGDVORTNYS
NALRNALMSTGSPNEFEVHEVDLIOMTSOEINVDVDSGPQYRRSSSGGGGCGO
GGPVYITLIVTVGGSGGQPPSGAGSGTGIAPTIADPSAGGVRSASPSGSG
PSGSRPSSSGSPGTRSPNGASGSSPGIAPGSGNSGAGVSGATGPASGSGYGP
STGATGSPGSGSPFPGVAGGSPYSPGAGPAGAGAGAGAGAGAGAGAGAGAGAG
GPGAGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GPGAGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GPGAGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GPGAGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
ACPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
IDADDPITITSELTITGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
FPGAGVPGSGSPGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG
SAGSAGATYPPGFGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
PCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GPGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GSGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GSGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
SVAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BASE COUNT 546 a 604 c 1087 g 593 t
ORIGIN
Query Match 3 8%; Score 86.6; DB 30; Length 2830;
Best Local Similarity 56.5%; Pred. No. 5.5e-09;
Matches 161; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
Qy 1519 gaggtcctgcagagagagagagagagagagagagagagagagagagagagagagagag 1578
Db 2107 GAGGACTTATGAGACCTGTGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2166
Qy 1579 gcagagagattccaaagccaggtgagagagagagagagagagagagagagagagagag 1638
Db 2167 CTGGTGAACCTTACGAGTGTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226
Qy 1639 gaccacaagcagaggttcgagagagagagagagagagagagagagagagagagagagag 1698
Db 2227 GACCTTACGAGACAGGTGTGTGTGTGACCTTACGAGAGAGAGAGAGAGAGAGAGAGAG 2286
Qy 1699 aacagcagaggttcgagagagagagagagagagagagagagagagagagagagagagag 1758
Db 2287 ACGGCGCGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2346
Qy 1759 caggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1803
Db 2347 TAGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2391

```

```

RESULT 11
AF218621
LOCUS AF218621 1499 bp DNA INV 26-FEB-2000
DEFINITION Nephila clavipes flagelliform silk protein (Flag) gene, upstream
partial cds.
ACCESSION AF218621
VERSION AF218621.1 GI:7106221
KEYWORDS
SEGMENT
1 of 2
Nephila clavipes.

```


Consensus quality: 190702 bases at least Q40

Consensus quality: 197814 bases at least Q30
Consensus quality: 202387 bases at least Q20
Insert size: 206666: sum-of-coverage
Insert size: 204428: 7.0% error: agarose-tp
Quality coverage: 3.36x in Q20 bases; sum-of-coverage
Quality coverage: 3.39x in Q20 bases; agarose-tp

NOTE: This is a 'working draft' sequence. It currently
consists of 27 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 17023: contig of 17023 bp in length
17024 17123: gap of 100 bp
17124 19850: contig of 2727 bp in length
19851 19950: gap of 100 bp
19951 24518: contig of 4568 bp in length
24519 24618: gap of 100 bp
24619 26755: contig of 2137 bp in length
26756 26855: gap of 100 bp
26856 28169: contig of 1314 bp in length
28170 28269: gap of 100 bp
28270 44370: contig of 16101 bp in length
44371 44470: gap of 100 bp
44471 53937: contig of 9467 bp in length
53938 54037: gap of 100 bp
54038 57906: contig of 3869 bp in length
57907 58006: gap of 100 bp
58007 69041: contig of 11035 bp in length
69042 69141: gap of 100 bp
69142 78379: contig of 9238 bp in length
78380 78479: gap of 100 bp
78480 84390: contig of 5911 bp in length
84391 84490: gap of 100 bp
84491 100259: contig of 15769 bp in length
100260 100359: gap of 100 bp
100360 110517: contig of 10158 bp in length
110518 110617: gap of 100 bp
110618 130776: contig of 20159 bp in length
130777 130876: gap of 100 bp
130877 132450: contig of 1574 bp in length
132451 132550: gap of 100 bp
132551 134957: contig of 2407 bp in length
134958 135057: gap of 100 bp
135058 142255: contig of 7198 bp in length
142256 142355: gap of 100 bp
142356 151246: contig of 8891 bp in length
151247 151346: gap of 100 bp
151347 160808: contig of 9462 bp in length
160809 160908: gap of 100 bp
160909 162264: contig of 1356 bp in length
162265 162364: gap of 100 bp
162365 164677: contig of 2313 bp in length
164678 164777: gap of 100 bp
164778 168335: contig of 3558 bp in length
168336 168435: gap of 100 bp
168436 170220: contig of 1785 bp in length
170221 170320: gap of 100 bp
170321 186878: contig of 16558 bp in length
186879 186978: gap of 100 bp
186979 194650: contig of 7672 bp in length
194651 194750: gap of 100 bp
194751 201357: contig of 6607 bp in length
201358 201457: gap of 100 bp
201458 209266: contig of 7809 bp in length.

FEATURES
Location/Qualifiers

1. 209266
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"

/map="q33.1-34"
/clone="RP11-88E10"
/clone_1ib="RP11-11.1"
1. 17023
/note="assembly_fragment:00123"
clone_end:17
vector_side:left"
17124. 19850
/note="assembly_fragment:00652"
fragment_chain:1"
19951. 24518
/note="assembly_fragment:00811"
fragment_chain:1"
24619. 26755
/note="assembly_fragment:00069"
fragment_chain:1"
26856. 28169
/note="assembly_fragment:01170"
fragment_chain:1"
28270. 44370
/note="assembly_fragment:00034"
fragment_chain:2"
44471. 53937
/note="assembly_fragment:00051"
fragment_chain:2"
54038. 57906
/note="assembly_fragment:01430"
fragment_chain:2"
58007. 69041
/note="assembly_fragment:00077"
fragment_chain:3"
69142. 78379
/note="assembly_fragment:01441"
fragment_chain:3"
78480. 84390
/note="assembly_fragment:00164"
fragment_chain:3"
84491. 100259
/note="assembly_fragment:00976"
fragment_chain:4"
100360. 110517
/note="assembly_fragment:01541"
fragment_chain:4"
110618. 130776
/note="assembly_fragment:00730"
fragment_chain:4"
130877. 132450
/note="assembly_fragment:00233"
fragment_chain:5"
132551. 134957
/note="assembly_fragment:01475"
fragment_chain:5"
135058. 142255
/note="assembly_fragment:00096"
142356. 151246
/note="assembly_fragment:00288"
151347. 160808
/note="assembly_fragment:00433"
160909. 162264
/note="assembly_fragment:00773"
162365. 164677
/note="assembly_fragment:00997"
164778. 168335
/note="assembly_fragment:01211"
168436. 170220
/note="assembly_fragment:01282"
170321. 186878
/note="assembly_fragment:01714"
186979. 194650
/note="assembly_fragment:01629"
fragment_chain:6"
194751. 201357
/note="assembly_fragment:01625"

```
misc-feature      201458..209266
                  /note="assembly-fragment:00434
                  /fragment_chain:6
                  /one_end:SP6
                  /vector_side:right"
BASE COUNT      57130 a 46237 c 46349 g 56891 t 2659 others
ORIGIN

Query Match      3.4%; Score 75.4; DB 66; Length 209266;
Best Local Similarity 59.6%; Pred. No. 2e-06;
Matches 127; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1552 ccagagagaccacaacagcaggttgggagagattcccaaaagcaggtgggagcgag 1611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70252 CTAGGGAACACACACACACACCAATAGGAAACACACACACACACACACACACAGG 70311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1612 gatccaaacagcaggttgggagagaccacaacagcaggtcgagcagagagaccca 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70312 AACACACACACACACACCACTAGGGAACACACACACACACACACACACACACAC 70371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1672 aacagcaggtcgagcagagagacccaacagcaggtcgagcagagagacccaacagc 1731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70372 ACCAGCAGGAGGCTACTAGGGAATACCAACACACACACACACACACACACACACAC 70431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1732 caggtcgagcagagagacccaacagcaggtg 1764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70432 CAGGACCTACTAGGGAACACACACACACACAGGG 70464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AF218623S1 14355 bp DNA INV 26-FEB-2000
LOCUS      AF218623S1
DEFINITION Nephila madagascariensis flagelliform silk protein (flag) gene,
            upstream partial cds.
ACCESSION  AF218623
VERSION     AF218623.1 GI:7106226
KEYWORDS
SEGMENT
SOURCE
ORGANISM   Nephila madagascariensis.
            Nephila madagascariensis
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
            Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
REFERENCE  1 (bases 1 to 14355)
            Hayashi,C.Y. and Lewis,R.V.
            Molecular architecture and evolution of a modular spider silk
            protein gene
            Science 287 (5457), 1477-1479 (2000)
JOURNAL    MEDLINE
MEDLINE    20156766
AUTHORS     2 (bases 1 to 14355)
            Hayashi,C.Y. and Lewis,R.V.
            Direct Submission
            Submitted (22-DEC-1999) Molecular Biology, University of Wyoming,
            P.O. Box 3944, Laramie, WY 82071, USA
FEATURES
SOURCE
location/qualifiers
1..14355
/organism="Nephila madagascariensis"
/db_xref="taxon:115969"
join(<1..845,3466..3864,4381..5739,7062..8078,9528..10814,12264..13481)
/gene="Flag"
/product="flagelliform silk protein"
<1..>14355
/gene="Flag"
join(473..845,3466..3864,4381..5739,7062..8078,9528..10814,12264..13481)
/gene="Flag"
/note="fibroin"
/codon_start=1
/product="flagelliform silk protein"
/protein_id="AF36091.1"
/db_xref="GI:7106228"
```

```
/translation="MGKGRHDTAKAKMOVALASSIAELVIAESSGQVORNTYIS
NALRNLMTSTGSPNEEFVHEVDLIOMLSODINPVDTSRPOQYRSPSSSGGGGG
GGPYITETLITVYVGGSGAGCAPSGAGPBGTCGYAPTGAPAGSGPGGVRPASPGSGG
PSGRSPSSSSSGSTRSPANKAGSGSPGILAPGSSSPGAGVSATGSPASSGSGYSGCT
TGATVGGGSEEPGPAAGGQYGPAGAGGAGAAVAPGVGVGGAGAPGVGGAGAP
GGYVGGAGPAGYGPAGAGPGYGPAGAGGGYGPAGAGPGYGPAGAGPGYGPAGS
GTGAGPGGYTPAGAGPGYGPAGYGPAGSGPGAGSGPGYGPAGYGPAGAGPGAGP
GAGPGAGPGAGPGAGCGATGTGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPG
RGAGRGAGRGAGRGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
PTITSELTITGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GIGSPGAGYGPAGSGSGSIPGSGSQQGILGPGSSQGGILGPGSSPGYSGGVGGP
YGRGSGSPGVGGAGGPTGPGSGGPGAGAGPYGPAGPGAGAGPYGPAGAGPGY
PGGYVGGAGPGAGPGAGAGPYGPAGPGAGPGYGPAGPGAGPGYGPAGAGPGAG
GPAGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG
GSGGATITIEDLITVDGANGPITISELTITGAGAGAGAGAGAGAGAGAGAGAGAG
GPGSGSGGAGYSGSGSPGSGSPGYSRPGYSGSRPDTGPGSGPGVGPAGVGP
YGRGSGSPGAGAGAGSYRPGYSGSGSPGAGAGPYGPAGAGPYGPAGPYGPAG
AGPGAGPGAGAGPYGPAGPGAGPGYGPAGAGPGYGPAGAGPGYGPAGAGPGAG
YGRGAGPGYGPAGPGAGPGAGPGAGPGYGPAGPGYGPAGPGAGPGAGPGAGPGAG
GSGPGYGPAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG
VGGAGSGGATITIEDLITVDGANGPITISELTITGAGAGAGAGAGAGAGAGAGAG
GPGYVGPAGSGPGYGPAGAGPYRPGSGSPGAGAGAGAGAGAGAGAGAGAGAGAG
YGRGSGSPGSGSPYGPAGPGAGPYGPAGAGPYGPAGAGPYGPAGAGPYGPAG
GGAGPYGPAGAGPYGPAGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGP
AGGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAG
AGGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAG
GGYGPAGTGPAGSGPGYGPAGSGPGYGPAGSGPGYGPAGSGPGYGPAGSGPGAG
GPGAGPGAGPYGPAGSGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG
SGGAGSGGATITIEDLITVDGANGPITISELTITGAGAGAGAGAGAGAGAGAGAG
GPGYVGPAGSGPGYGPAGAGPYRPGSGSPGAGAGPGAGAGPYGPAGAGPYGPAG
YGRGSGPGAGAGPYGPAGPGAGPYGPAGPGAGPYGPAGAGPYGPAGAGPYGPAG
GGAGPYGPAGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGPAGPGAGPYGP
```

```
BASE COUNT      3976 a 2647 c 3677 g 4051 t 4 others
ORIGIN

Query Match      3.3%; Score 75; DB 31; Length 14355;
Best Local Similarity 59.7%; Pred. No. 2.3e-06;
Matches 126; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1557 ggaactcaaacagcaggttgggagagattccaaagcaggttgggagcgagatcc 1616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10565 GGAAGTGAAGAGACCCGCTGTGCTGACCTACGACCTGAGAGTCCAGGTGACCC 10624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1617 aaacagcaggttgggagagaccacaacagcaggtcgagcagagagacccaacag 1676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10625 TACGACCGGGTGTGACAGTGTGACCTACGACAGGTGTGACAGTACCTACGGA 10684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1677 ccaagtcgagcagagagacccaacagcaggttcgagcagagagacccaacagcaggt 1736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10685 CAGAGTGAAGAGAGTGAACCTACGACCGGGTGTGCTGTGACCTACGACCGGTT 10744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1737 cgggccaagagacccaacagcaggttggt 1767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10745 GGTGCTGTGTGACCTACGACCTGTGAGGTG 10775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: October 14, 2000, 05:00:16
Job time: 8208 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 17:28:24 ; Search time 262.03 Seconds
(without alignments)
1321.191 Million cell updates/sec

Title: US-09-431-843B-11

Perfect score: 2289
Sequence: 1 tagaatcagcgagcgctga.....cctctgagtgaggcgcg 2289

Scoring table: IDENTITY_NIC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PCtUS_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	76	3.3	7218	1	US-08-232-463-14 Sequence 14, Appl
C 2	75.8	3.3	6192	3	US-08-479-537A-1 Sequence 1, Appl
C 3	75.8	3.3	6449	3	US-08-479-537A-4 Sequence 4, Appl
C 4	74.6	3.3	1931	4	US-08-130-114-2 Sequence 2, Appl
C 5	67.2	2.9	2580	5	US-09-050-863-2 Sequence 2, Appl
C 6	67.2	2.9	5452	4	US-09-130-114-1 Sequence 1, Appl
C 7	67.2	2.9	10596	1	US-07-884-811-15 Sequence 15, Appl
C 8	67.2	2.9	10596	1	US-07-885-971-15 Sequence 15, Appl
C 9	67.2	2.9	10596	1	US-08-087-783A-15 Sequence 15, Appl
C 10	67.2	2.9	10596	1	US-08-194-088B-15 Sequence 15, Appl
C 11	67.2	2.9	10596	3	US-08-194-088B-15 Sequence 15, Appl
C 12	67.2	2.9	10596	6	PCT-US93-04648-15 Sequence 15, Appl
C 13	65.8	2.9	35100	3	US-08-739-379-17 Sequence 17, Appl
C 14	61.8	2.7	8438	1	US-07-945-283-1 Sequence 1, Appl
C 15	60.2	2.6	3489	4	US-08-728-323A-1 Sequence 1, Appl
C 16	60.2	2.6	32207	3	US-08-770-379-20 Sequence 20, Appl
C 17	58.4	2.6	130	3	US-07-829-461A-8 Sequence 8, Appl
C 18	58.4	2.6	51259	5	US-08-781-891-209 Sequence 209, App
C 19	56.2	2.5	2277	2	US-08-676-967-5 Sequence 5, Appl
C 20	56.2	2.5	2277	4	US-08-676-974-5 Sequence 5, Appl
C 21	56.2	2.5	2277	4	US-09-098-487-5 Sequence 5, Appl
C 22	56	2.4	16442	5	US-08-781-891-208 Sequence 208, App
C 23	54.8	2.4	768	5	US-08-946-026-32 Sequence 32, Appl
C 24	54	2.4	711	2	US-08-452-531-2 Sequence 2, Appl
C 25	54	2.4	711	2	US-08-460-746A-2 Sequence 2, Appl
C 26	54	2.4	711	2	US-08-460-555-2 Sequence 2, Appl

C 27	54	2.4	1091	2	US-08-452-531-1 Sequence 1, Appl
C 28	54	2.4	1091	2	US-08-460-746A-1 Sequence 1, Appl
C 29	54	2.4	1091	2	US-08-460-555-1 Sequence 1, Appl
C 30	50.8	2.2	2277	2	US-08-676-967-2 Sequence 2, Appl
C 31	50.8	2.2	2277	2	US-08-676-974-2 Sequence 2, Appl
C 32	50.8	2.2	2277	4	US-09-098-487-2 Sequence 2, Appl
C 33	50.8	2.2	2301	2	US-08-306-691B-23 Sequence 23, Appl
C 34	50.8	2.2	2301	6	PCT-US93-06251-78 Sequence 78, Appl
C 35	50.6	2.2	53526	5	US-08-658-136-2 Sequence 2, Appl
C 36	50.6	2.2	53577	5	US-08-658-136-2 Sequence 2, Appl
C 37	49.8	2.2	2824	4	US-09-010-928B-3 Sequence 3, Appl
C 38	48	2.1	1052	2	US-08-466-603-1 Sequence 1, Appl
C 39	48	2.1	1052	2	US-08-314-503A-1 Sequence 1, Appl
C 40	48	2.1	1052	2	US-08-468-066-1 Sequence 1, Appl
C 41	48	2.1	1052	3	US-08-466-717-1 Sequence 1, Appl
C 42	48	2.1	1052	5	US-08-466-743-1 Sequence 1, Appl
C 43	48	2.1	1052	6	PCT-US95-12414-1 Sequence 1, Appl
C 44	47.4	2.1	496	2	US-08-263-413-23 Sequence 23, Appl
C 45	47.4	2.1	500	2	US-08-263-413-22 Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTgpt-F15
US-08-232-463-14


```

Matches 113; Conservative 0; Mismatches 64; Indels 0; Gaps 0
QY 10 ggcggccgtgaattcttagcgcagcatgagacgaccgccactgcactccaacctgagagag 69
Db 1003 GTGACCGAGAGACTGGGGCCGCAGGTGAGAGACGAGACGCCGAGAGACGAGCACTGGGAG 944
QY 70 gagagagagatgctcgaaaggacgcgcggagagacgaagtctcgaaagagcgagagccgcgc 129
Db 943 GAGGACGAGAGACCGGGAGGACGGGGGAGGACGAGACGGGAGGACGGGAGGACGAGGAC 884
QY 130 gcgaaggagcacgcagcagggagagagagcgagagtcggaagagacgcgcggcgagcg 186
Db 883 GGCGAGAGACGGGAGGAGACGAGCAGCGGGGAGGAGACGGGAGAGAGACGAGACGGGAG 827

RESULT 5
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
GENERAL INFORMATION:
APPLICANT: lao, ying
APPLICANT: Hiang, Betty
APPLICANT: Payan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050, 863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 2.9%; Score 67.2; DB 5; Length 2580;
Best Local Similarity 47.7%; Pred. No. 2,le-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;
QY 956 ggaaggttgagagagaagaagccccgcggggagacccgcagcaagagccagcacccaggtc 1015
Db 707 GCACGAGACCAGAGGAGGAGGCGCAGAGGCGCACGAGAGGCGCAGAGGCGCAGAGAGGAGG 766
QY 1016 ggaacctgtgggcacaagcatagcaaaggttgr---ggcgaggttgagcagagggcccacg 1072
Db 767 AGGGCGCAGAGCAGGAGAGGAGCGCCGAGAGGGCGCAGAGAGGGCGCAGAGACGACGAGGGCC 826
QY 1073 caaggagcgttgagagcccccagagatgcggagaccccttgagagagagcaccaggggatgagcg 1132

```

[illegible]

Qy	1073	cacggagaggttgagagcccccagatctcggagaccctctgagagagagccaaagggagtgagcg	1132
Db	2407	AGGACACAGAGAGAGGGCGAGAGGGCGAGACAGAGAGAGGGCGAGAGGGCGAGAGG	2466
Qy	1133	ggggccacgggggaagatagcgcggagcccttaagccccaagagagcaagaagaggaagc	1192
Db	2467	GGCAGGACACAGAGAGAGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2526
Qy	1193	tggagctgagccggtcgggagagagccgcccaagaccagagccctcagagtgctcctaaggg	1252
Db	2527	AGGGCAGAGAGGGCGAGAGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2586
Qy	1253	tggagaagatcgctctgaatttggagaggtgtgtccctcagcaagggcagccctcaggaagc	1312
Db	2587	AGGGCAGAGAGGGCGAG	2646
Qy	1313	ggaccaggaaggttggcggtcaggaacctgtgggagcgagtgcaagccctgcggcaacccc	1372
Db	2647	AGGGCAGAGACAGAGAGGGCGAGAGCGAGAGCGCGAGAGAGAGAGAGAGGGCGAGAGCAG	2706
Qy	1373	tggagagcaaggttggcgcacaaggtgagagagcgaggaaggttgataggtgtctgggg	1432
Db	2707	AGGGCAGAGAGGGCGAG	2766
RESULT	9		
US-08-087-783A-15			
Sequence 15, Application US/08087783A			
Patent No. 5547856			
GENERAL INFORMATION:			
APPLICANT: Godswalk, Paul J., Lokker, Nathalie A., Mark, Melanie R.			
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS			
NUMBER OF SEQUENCES: 22			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Genentech, Inc.			
STREET: 460 Point San Bruno Blvd			
CITY: South San Francisco			
STATE: California			
COUNTRY: USA			
ZIP: 94080			
COMPUTER READABLE FORM:			
MEDIUM TYPE: 3.5 Inch, 1.44 Mb floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Winpatin (Genentech)			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/087,783A			
FILING DATE: 13-Jul-1993			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 07/884811			
FILING DATE: 18-MAY-92			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 07/885971			
FILING DATE: 18-MAY-92			
ATTORNEY/AGENT INFORMATION:			
NAME: Marschang, Diane L.			
REGISTRATION NUMBER: 35,600			
REFERENCE/DOCKET NUMBER: P0755779P1			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: 415/225-5416			
TELEFAX: 415/952-9881			
TELEX: 910/371-7168			
INFORMATION FOR SEQ. ID NO: 15:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 10596 base pairs			
TYPE: Nucleic Acid			
STRANDEDNESS: Single			
TOPOLOGY: Linear			
US-08-087-783A-15			
Query Match	2.9%	Score 67.2	DB 1; Length 10596;

```

Best Local Similarity 47.7%; Pred. No. 2,8e+05; Indels 3; Gaps 1;
Matches 229; Conservative 0; Mismatches 246;
?
QY 956 ggaaggtggaaggaaggaagcccccggagaccaccgaagaccgaaccgaaggtc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2287 GGGAGAGAGCAGAGGAGGGGGCGAGAGGGGCCAGAGAGGGGCGACGAGCGAGG 2346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1016 ggaacctgtggccaaagacataagcaaggttgg--ggcaaggttgaaagagggcccaagc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2347 AGGGGACGAGAGCAGAGAGGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGGGGCG 2406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1073 caacgagcgttgaagcccccaagatctcgggaaaccccttgaaagagaccagggggatgaagcag 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2407 AGAGCAGAGAGGAGAGGGGCGAGAGGGCGAGAGCAGAGAGAGGCGCAGAGAGGCGCAGAGG 2466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1133 gggggccacggggaaagatgtgcggggcccttaagccccaagaagagacaaagaagaaagc 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2467 GCGAGGAGCAGAGAGAGGGGCGAGAGCGAGAGAGAGGGGCGAGAGGCGCAGAGAGCGAGGAGG 2536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1193 ttgagctgagccggcgggagccagccgcccacaaagacaaagcccttaagatgtcctcaagag 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2527 AGGGGCGAGGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGAGGGCGAGAGAGGCGCAGG 2586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1253 ttgagaagaatcgctcttgaatttgaaaggtgtgtcccttaagccaaagcgagcctcaggaagc 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2587 AGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCAGAGAGAGGCGCAGAGAGGGGCGAGG 2646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1313 ggaaccagaagttggcggttcagagcccttgggaagcagttcaagccctgcggccaaacccc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2647 AGGGGCGAGAGAGGAGGGGCGAGAGCGAGAGAGGGGCGAGAGAGGCGCAGAGAGCGAGG 2706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1373 tgggaagccaaaggttggcgcagacaaggttgaagaaagcggaaaggttgatgtaggtgtctgggg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 AGGGCGAGAGGAGGGCGAGAGCGAGAGGGGCGAGAGGGGCGAGAGAGGCGCAGAGAGG 2766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Marf, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCTITE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 75SD1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
;
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-088b-15

Query Match 2.9% Score 67.2; DB 1; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

OY 956 ggaaggtgaggaaggaagccccgggagccccgaccagagccagccaggggtc 1015
DB 2287 GGCAGGACAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2346
OY 1016 ggaacctgtggccagagcatagcaaggtgtg---ggcagaggtgagcagagggcccaagc 1072
DB 2347 AGGGCAGAGAGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2406
OY 1073 caccgagcgtggagcccccagagatgcggagacccttgagagagagcagagggatgagcag 1132
DB 2407 AGGAGCAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2466
OY 1133 ggggcaagggggaagatagcccgagcccttaagccccaaagagagcaagaggaagc 1192
DB 2467 GGCAGGACAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2526
OY 1193 tggagctagccgagcgagagcagccgcccacagagccagagcccttaagatgctcctaagag 1252
DB 2527 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2586
OY 1253 tggagaagaatcgctctgaatttgagaggggtgtgtccctcaagggcgagccctcaagagc 1312
DB 2587 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2646
OY 1313 ggaaccagaaggtggcggttcagagacccttgaggagcagtgcaagccctgcgcacacccc 1372
DB 2647 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2706
OY 1373 tggagcagaaggtggcgagcaaggtgaggaagcggaggaaggtgagtgaggtgtctggag 1432
DB 2707 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2766

RESULT 11

US-08-194-087-15
Sequence 15, Application US/08194087
GENERAL INFORMATION:
PATENT NO. 5879910
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEINASE DOMAIN VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194.087
CLASSIFICATION: 435
FILING DATE: 18-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 779
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-087-15

Query Match 2.9% Score 67.2; DB 3; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

OY 956 ggaaggtgaggaaggaagccccgggagccccgaccagagccagccaggggtc 1015
DB 2287 GGCAGGACAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2346
OY 1016 ggaacctgtggccagagcatagcaaggtgtg---ggcagaggtgagcagagggcccaagc 1072
DB 2347 AGGGCAGAGAGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2406
OY 1073 caccgagcgtggagcccccagagatgcggagacccttgagagagagcagcagggatgagcag 1132
DB 2407 AGGAGCAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2466
OY 1133 ggggcaagggggaagatagcccgagcccttaagccccaaagagagcaagaggaagc 1192
DB 2467 GGCAGGACAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2526
OY 1193 tggagctagccgagcgagagcagccgcccacagagccagagcccttaagatgctcctaagag 1252
DB 2527 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2586
OY 1253 tggagaagaatcgctctgaatttgagaggggtgtgtccctcaagggcgagccctcaagagc 1312
DB 2587 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2646
OY 1313 ggaaccagaaggtggcggttcagagacccttgaggagcagtgcaagccctgcgcacacccc 1372
DB 2647 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2706
OY 1373 tggagcagaaggtggcgagcaaggtgaggaagcggaggaaggtgagtgaggtgtctggag 1432
DB 2707 AGGGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGG 2766

RESULT 12

PCT-US93-04648-15
Sequence 15, Application PCT/US9304648
GENERAL INFORMATION:
PATENT NO. 5879910
APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648


```

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: Curtis P. Ribando
STREET: 1815 No. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext.513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")

```

[illegible]

```

US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 17:09:44 ; Search time 262.03 Seconds

(without alignments)
1389.877 Million cell updates/sec

Title: US-09-431-843b-5

Perfect score: 2408
Sequence: 1 tagaattcggcgccgcgtga.....ccctctggtggtggtggcgcgcg 2408

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5C.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/6.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/PCUS.COMB.seq:*
7: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	4.2	6192	3 US-08-479-537A-1	Sequence 1, Appl
2	101	4.2	6449	3 US-08-479-537A-4	Sequence 4, Appl
3	76	3.2	7218	1 US-08-232-463-14	Sequence 14, Appl
4	74.6	3.1	1931	4 US-09-130-114-2	Sequence 2, Appl
5	68.2	2.8	35100	3 US-08-770-379-17	Sequence 17, Appl
6	67.2	2.8	2580	5 US-09-050-863-2	Sequence 2, Appl
7	67.2	2.8	5452	4 US-09-130-114-1	Sequence 1, Appl
8	67.2	2.8	10596	1 US-07-884-811-15	Sequence 15, Appl
9	67.2	2.8	10596	1 US-07-885-971-15	Sequence 15, Appl
10	67.2	2.8	10596	1 US-08-087-783A-15	Sequence 15, Appl
11	67.2	2.8	10596	1 US-08-194-088B-15	Sequence 15, Appl
12	67.2	2.8	10596	6 PCT-US93-04648-15	Sequence 15, Appl
13	61.8	2.6	8438	1 US-07-945-283-1	Sequence 1, Appl
14	60.2	2.5	3489	4 US-08-728-323A-1	Sequence 1, Appl
15	60.2	2.5	32207	3 US-08-770-379-20	Sequence 20, Appl
16	58.8	2.4	711	2 US-08-452-531-2	Sequence 2, Appl
17	58.8	2.4	711	2 US-08-460-746A-2	Sequence 2, Appl
18	58.8	2.4	711	2 US-08-460-555-2	Sequence 2, Appl
19	58.8	2.4	1091	2 US-08-452-531-1	Sequence 1, Appl
20	58.8	2.4	1091	2 US-08-460-746A-1	Sequence 1, Appl
21	58.8	2.4	1091	2 US-08-460-555-1	Sequence 1, Appl
22	58.8	2.4	150	3 US-07-829-461A-8	Sequence 8, Appl
23	58.4	2.4	51259	5 US-08-781-891-109	Sequence 109, Appl
24	56.2	2.3	2277	2 US-08-676-967-5	Sequence 5, Appl
25	56.2	2.3	2277	2 US-08-676-967-5	Sequence 5, Appl
26	56.2	2.3	2277	2 US-08-676-974-5	Sequence 5, Appl

ALIGNMENTS

27	56.2	2.3	2277	4 US-09-098-487-5	Sequence 5, Appl
28	56	2.3	16442	5 US-08-781-891-208	Sequence 208, App
29	54.8	2.3	7668	5 US-08-946-026-32	Sequence 32, Appl
30	53	2.2	2824	4 US-09-010-928B-3	Sequence 3, Appl
31	53	2.2	6192	3 US-08-479-537A-1	Sequence 1, Appl
32	53	2.2	6449	3 US-08-479-537A-4	Sequence 4, Appl
33	50.8	2.1	2277	2 US-08-676-967-2	Sequence 2, Appl
34	50.8	2.1	2277	2 US-08-676-974-2	Sequence 2, Appl
35	50.8	2.1	2277	4 US-09-098-487-2	Sequence 2, Appl
36	50.8	2.1	2301	2 US-08-306-691B-23	Sequence 23, Appl
37	50.8	2.1	2301	6 PCT-US93-06251-78	Sequence 78, Appl
38	50.6	2.1	53526	5 US-08-658-136-2	Sequence 2, Appl
39	50.6	2.1	53577	5 US-08-658-136-1	Sequence 1, Appl
40	48	2.0	1052	2 US-08-466-603-1	Sequence 1, Appl
41	48	2.0	1052	2 US-08-314-503A-1	Sequence 1, Appl
42	48	2.0	1052	2 US-08-468-066-1	Sequence 1, Appl
43	48	2.0	1052	3 US-08-466-717-1	Sequence 1, Appl
44	48	2.0	1052	5 US-08-466-743-1	Sequence 1, Appl
45	48	2.0	1052	6 PCT-US95-12414-1	Sequence 1, Appl

RESULT 1
US-08-479-537A-1
Sequence 1, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMBER, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note="The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note="Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note="Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note="Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-1

Query Match 4.2% Score 101; DB 3; Length 6192;
Best Local Similarity 45.4%; Pred. No. 2e-12;

Matches 215; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

OY 1585 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 1644
DB 461 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 520
OY 1645 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 1704
DB 521 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 580
OY 1705 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 1764
DB 581 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 640
OY 1765 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 1824
DB 641 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 700
OY 1825 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 1884
DB 701 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 760
OY 1885 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 1944
DB 761 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 820
OY 1945 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 2004
DB 821 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNNG 880
OY 2005 ccacggagagaccacggcccgccagcagacctgcagggagagcagcagcagacc 2058

DB 881 CCCAGGGTGTCACTCGGCCCCGGGACNNNAGCCGNNNCCGGGCTCCACCGCCCC 934
RESULT 2
US-08-479-537A-4
Sequence 4, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L. 35,030
REGISTRATION NUMBER: 017753-025
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SRO ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note="The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..5661
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457


```

? OTHER INFORMATION: /note="Nucleotide 457 is X1 = NNN
? OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
? OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 487
? OTHER INFORMATION: /note="Nucleotide 487 is Y = NNN
? OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
? OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 496
? OTHER INFORMATION: /note="Nucleotide 496 is X2 = NNN
? OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
? OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
? OS-08-479-537A-4

```

```

Query Match Similarity      4.2%      Score 101; DB 3;  Length 6449;
Best Local Similarity      45.4%      Pred. No. 2e-12;
Matches 215; Conservative  0;  Mismatches 259;  Indels  0;  Gaps  0;

```

QY	1885	ccatcggagagcccagagcccagcccagcagcagacgtcgcaggggagagagcagcagaagac	16444
Db	461	CCACGGGTGTCACTCGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	520
QY	1645	ccatcggagagcccagagcccagcccagcccagcagagactacaagggatgagccagccgagac	17043
Db	521	CCCAAGGTGTCACTCGGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	580
QY	1705	ccatcggagagcccagagcccagcccagcccagcagagactcgcaggggagcagccagccgagac	17643
Db	581	CCCAAGGTGTCACTCGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	640
QY	1765	ccatcggagagcccagagcccagcccagcccagcagactcgcagggagcagccagccagactc	18243
Db	641	CCCAAGGTGTCACTCGGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	700
QY	1825	ccatcggagagcccagagcccagcccagcccagcagagctacaagggatgagccagccgagagc	18843
Db	701	CCCAAGGTGTCACTCGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	760
QY	1885	ccatcggagagcccagagcccagcccagcccagcagagctcgcaggggagcagccagccgagac	19444
Db	761	CCCAAGGTGTCACTCGGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	820
QY	1945	ccatcggagagcccagagcccagcccagcccagcagagctcgcaggggagcagccagccgagac	20044
Db	821	CCCAAGGTGTCACTCGGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCCC	880
QY	2005	ccatcggagagcccagagcccagcccagcccagcagagctacaagggatgagccagccagcc	2058
Db	881	CCCAAGGTGTCACTCGGGCCCCCGGACNNNAGCGCGNNCCGGGGCTTCACCGCCCC	934

```

1  RESULT
2  US-08-232-463-14/c
3  : Sequence 14, Application US/08232463
4  : Patent No. 5670367
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: DORNER, F.
9  : APPLICANT: SCHEIFLINER, F.
10 :
11 : APPLICANT: FALKNER, F. G.
12 :
13 : TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
14 :
15 : NUMBER OF SEQUENCES: 52
16 :
17 : CORRESPONDENCE ADDRESSES:
18 :
19 : ADDRESSEE: Foley & Lardner
20 :
21 : STREET: 1800 Diagonal Road, Suite 500
22 :
23 : City: Alexandria
24 :
25 : STATE: VA
26 :
27 : COUNTRY: USA
28 :
29 : ZIP: 22313-0299
30 :
31 : COMPUTER READABLE FORM:

```

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/232,463
7 FILING DATE:
8 CLASSIFICATION: 435
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US/07/935,313
11 FILING DATE:
12 APPLICATION NUMBER: EP 91 114 300.6
13 FILING DATE: 26-AUG-1991
14 ATTORNEY/AGENT INFORMATION:
15 NAME: BENT, Stephen A.
16 REGISTRATION NUMBER: 29,768
17 REFERENCE/DOCKET NUMBER: 30477/114 IMMU
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (703)836-9300
20 TELEFAX: (703)683-4109
21 TELEX: 899149
22 INFORMATION FOR SEQ ID NO: 14:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 7218 base pairs
25 TYPE: nucleic acid
26 STRANDEDNESS: single
27 TOPOLOGY: linear
28 IMMEDIATE SOURCE:
29 CLONE: pTZ9pt-F15
30
31 US-08-232-463-14

```

Query Match	3.2%	Score 76	DB 1	Length 7218
Best Local	6.0%	Pred. No.	2.8e-07	
Matches	25	Conservative	155	Indels 0
		Mismatches	155	Gaps 0

[illegible]

RESULT 4
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damej, Bassam B.
; APPLICANT: Robbins, Alan K.

TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 2.8%; Score 67.2; DB 5; Length 2580;
Best Local Similarity 47.7%; Pred. No. 1.4e-05;

Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaagctggaaggaaggaagccccgggagaccacagagccagccagccagggc 1015
DB 707 ggacagacag 766
QY 1016 ggaagctggaagcagagcagagcagaggtg---ggcagagtggaagagggccagc 1072
DB 767 agggcagagcagagcagagagagagagagagagagagagagagagagagag 826
QY 1073 cagcagagcgtggaagccagagatgagcccttggaagagagagagagagagag 1132
DB 827 aggagcag 886
QY 1133 gggagcagaggaagatagccgagcccttaagcccaagagagagagagagag 1192
DB 887 ggacagacag 946
QY 1193 tggagctgagccgagagagagagagagagagagagagagagagagagagag 1252
DB 947 agggcag 1006
QY 1253 tggagagatcgctctgaatttgagaggtgtgtccctcagccagagcagagc 1312
DB 1007 agggcag 1066
QY 1313 ggaaccaggaagtggcggtcaggaacctgggagagcagtgccagccctgcgaaccc 1372
DB 1067 agggcag 1126
QY 1373 tggagcagaggtggtgcagcaaggtggaagagcagagagagagagagagag 1432
DB 1127 agggcag 1186

RESULT 7

US-09-130-114-1/C
Sequence 1, Application US/09130114
Patent No. 5976807
GENERAL INFORMATION:
APPLICANT: Horlick, Robert A.
APPLICANT: Damsi, Basam B.
APPLICANT: Robbins, Alan K.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
FILE REFERENCE: 0867/ID903US1
CURRENT APPLICATION NUMBER: US/09/130,114
CURRENT FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 5452
TYPE: DNA
ORGANISM: VEBNA
US-09-130-114-1

Query Match 2.8%; Score 67.2; DB 4; Length 5452;
Best Local Similarity 47.7%; Pred. No. 1.7e-05;

Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaagctggaaggaaggaagccccgggagaccacagagccagccagccagggc 1015
DB 2098 ggacagacag 2039
QY 1016 ggaagctggaagcagagcagagcagaggtg---ggcagagtggaagagggccagc 1072
DB 2038 agggcagagcag 1979
QY 1073 cagcagagcgtggaagccagagatgagcccttggaagagagagagagagagag 1132
DB 1978 aggagcag 1919
QY 1133 gggagcagaggaagatagccgagagcccttaagcccaagagagagagagagag 1192
DB 1918 ggacagacag 1859
QY 1193 tggagctgagccgagagagagagagagagagagagagagagagagagagag 1252
DB 1858 agggcag 1799
QY 1253 tggagagatcgctctgaatttgagaggtgtgtccctcagccagagcagagc 1312
DB 1798 agggcag 1739
QY 1313 ggaaccaggaagtggcggtcaggaacctgggagagcagtgccagccctgcgaaccc 1372
DB 1738 agggcag 1679
QY 1373 tggagcagaggtggtgcagcaaggtggaagagcagagagagagagagagag 1432
DB 1678 agggcag 1619

RESULT 8

US-07-884-811-15
Sequence 15, Application US/07884811
Patent No. 5316921
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/884,811
FILING DATE: 19920518
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ganger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases

Patent No. 5879910
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,087
FILING DATE: 18-MAY-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 779
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-087-15

Query Match 2.8%; Score 67.2; DB 3; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;
QY 956 ggaaggtggaaggaaggaagcccccgggagcccgaccagagccacccaggtc 1015
DB 2287 GGCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2346
QY 1016 ggaactgtgagccagagcatagcaaggtgtg--ggcaggtgtgacgggccccagc 1072
DB 2347 AGGGGACAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2406
QY 1073 caaggagcgtgagccccagagatcgcggaaccccttgagagagagcaggggagtgaagcag 1132
DB 2407 AGGACAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2466
QY 1133 gggagccaggggaagataagcgcggaagcccttaagccccaagagagcagaagagaagc 1192
DB 2467 GGCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2526
QY 1193 tggagctgagccggcgagagcagccgcccacagggcgcaagcccttaagatgtcctcaaggg 1252
DB 2527 AGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2586
QY 1253 tggagagatgcctctgaatttggaggggtgtgtccctcaagccagggcgagcctcaagagcgg 1312
DB 2587 AGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2646
QY 1313 ggaaccaggaagtggcggtcaggaaccccttgaggagagcaggtgaccccgccagccacccc 1372
DB 2647 AGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2706

QY 1373 tggagccaggggtgcccagacaagtgtgaggaagcggaagaggtgtgattgactgggg 1432
DB 2707 AGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2766
RESULT 13
PCT-US93-04648-15
Sequence 15, Application PC/TUS9304648
GENERAL INFORMATION:
APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melani
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648
FILING DATE: 19930517
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885971
FILING DATE: 18-MAY-92
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755,779P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US93-04648-15
Query Match 2.8%; Score 67.2; DB 6; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;
QY 956 ggaaggtggaaggaaggaagcccccgggagcccgaccagagccacccaggtc 1015
DB 2287 GGCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2246
QY 1016 ggaactgtgagccagagcatagcaaggtgtg--ggcaggtgtgacgggccccagc 1072
DB 2347 AGGGGACAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2406
QY 1073 caaggagcgtgagccccagagatcgcggaaccccttgagagagagcaggggagtgaagcag 1132
DB 2407 AGGACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2466
QY 1133 gggagccaggggaagataagcgcggaagcccttaagccccaagagagcagaagagaagc 1192
DB 2467 GGCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2526
QY 1193 tggagctgagccggcgagagcagccgcccacagggcgagcctctcaagatgtcctcaaggg 1252

LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-1

Query Match 3.8%; Score 88.4; DB 3; Length 6192;
Best Local Similarity 45.4%; Pred. No. 1.3e-09;

Matches 188; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 1585 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1644
DB 461 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 520
QY 1645 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1704
DB 521 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 580
QY 1705 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1764
DB 581 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 640
QY 1765 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1824
DB 641 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 700
QY 1825 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1884
DB 701 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 760
QY 1885 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1944
DB 761 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 820
QY 1945 ccacggagagcccccagagcccccagagccctcagagagcagcagccgagagc 1998
DB 821 ccacggagagcccccagagcccccagagcccccagagcccccagagcccccagagc 874

RESULT 2

US-08-479-537A-4
Sequence 4, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIEN, Marie-Paule
APPLICANT: LATHÉ, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEO ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fi
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..5661
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:

LENGTH: 1931
TYPE: DNA
ORGANISM: EBNA
US-09-130-114-2

Query Match 3.2% Score 74.6; DB 4; Length 1931;
Best Local Similarity 63.8%; Pred. No. 6.4e-07;
Matches 113; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 10 ggcggcgtatcttctagccgagcatgtgagaccccgactgctccactgtgagagag 69
DB 1003 GTGACGGAGAGCTGGGGCCGAGGTGGAGACGAGACGGGAGACGAGACGGGGAG 944
QY 70 gaagagagagatgtcgagagcgcgagagagagactgtcgagagcgagcgccgagc 129
DB 943 GAGGACGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 884
QY 130 ggcagagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 186
DB 883 GGGGAGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 827

RESULT 5
US-08-770-379-17
Sequence 17, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770.379
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-17

Query Match 3.0% Score 70.2; DB 3; Length 35100;
Best Local Similarity 48.8%; Pred. No. 8.9e-06;
Matches 221; Conservative 0; Mismatches 228; Indels 4; Gaps 1;

QY 1587 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 1646
DB 24281 AACGAGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGCA 24340
QY 1647 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 1706
DB 24341 CCCGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGCA 24400
QY 1707 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 1766
DB 24401 CCCGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGCA 24460
QY 1767 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 1826
DB 24461 CCCGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGCA 24520
QY 1827 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 1886
DB 24521 CCCGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGCA 24580
QY 1887 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 1946
DB 24581 CCCGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGAT 24640
QY 1947 atcgagagaccccccagccccagcagcagcagcagcagcagcagcagcagcagcagc 2006
DB 24641 CCCGAGAGCCCGGACAGACCCGAGAGCCCGGACAGACCCGAGAGCCCGGAGG 24696
QY 2007 ggaagcagcagagttgagcagcagcagcagcagcagcagcagcagcagcagcagc 2039
DB 24697 GGATCCGGCGCGCCGACACCTCCCGGAGGGGA 24729

RESULT 6
US-09-050-863-2
Sequence 2, Application US/09050863
Patent No. 611411
GENERAL INFORMATION:

APPLICANT: Lao, Ying
APPLICANT: Hiang, Betty
APPLICANT: Payan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

ADDRESS: Fiehn, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050.863
FILING DATE: 30-MAR-1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 2.9%; Score 67.2; DB 5; Length 2580;
Best Local Similarity 47.7%; Pred. No. 2.1e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggttgagaggaaggaagcccccgggagcccccgaacagagcgagaccagaggtc 1015
DB 707 ggcagagcagcagagggcagagggcagagggcagagggcagagggcagagggcagagggc 766
QY 1016 ggaacccgtgagcagagcagcagaggtg---ggcagaggtgagcagagggcagcagc 1072
DB 767 AGGGGAG 826
QY 1073 cagcagcgttgagagccagagatcgagagcccttgagagagagcagaggggagtgagcag 1132
DB 827 AGGAGCAG 886
QY 1133 gggagcagcagggagagagtaagcccgagcccttaagccccaagagagcagagagagc 1192
DB 887 GGCAGAGCAGCAG 946
QY 1193 tggagctgagccgagcagcagcagcccaagagcagccctcagagtgctcagagag 1252
DB 947 AGGGGAG 1006
QY 1253 tggagagagatcgctcgtgaatttgagaggtgtgctcctcagcagagcgagcctcagagcgg 1312
DB 1007 AGGGGAG 1066
QY 1313 ggaacccaggaagtgagcagcagcagcccttgagagcagtgagccctcagcagcagccccc 1372
DB 1067 AGGGGAG 1126
QY 1373 tggagcagcaggtgagcagcagagagtgagagagcagagagtgagtgagtgagtgagtg 1432
DB 1127 AGGGGAG 1186

RESULT 7
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Dama, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903051
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VERBNA
US-09-130-114-1

Query Match 2.9%; Score 67.2; DB 4; Length 5452;
Best Local Similarity 47.7%; Pred. No. 2.5e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggttgagaggaaggaagcccccgggagcccccgaacagagcgagaccagaggtc 1015
DB 2098 GGCAGAGCAGCAG 2039
QY 1016 ggaacccgtgagcagagcagcagaggtg---ggcagaggtgagcagagggcagcagcagc 1072

DB 2038 AGGGGAG 1979
QY 1073 cagcagcgttgagagcccccggagtgagagcccttgagagagagcagagggagtgagcag 1132
DB 1978 AGGAGCAG 1919
QY 1133 gggagcagcaggaagagtaagcccgagcccttaagccccaagagagcagagagagc 1192
DB 1918 GGCAGAGCAG 1859
QY 1193 tggagctgagccgagcagcagcagcccaagagcagccctcagagtgctcagagag 1252
DB 1858 AGGGGAG 1799
QY 1253 tggagagagatcgctcgtgaatttgagaggtgtgctcctcagcagagcgagccctcagagcgg 1312
DB 1798 AGGGGAG 1739
QY 1313 ggaacccaggaagtgagcagcagcagcccttgagagcagtgagccctcagcagcagccccc 1372
DB 1738 AGGGGAG 1679
QY 1373 tggagcagcaggtgagcagcagagtgagagagcagagagtgagtgagtgagtgagtg 1432
DB 1678 AGGGGAG 1619

RESULT 8
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCTYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-72168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-884-811-15

Query Match 2.9%; Score 67.2; DB 1; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;


```

: ATTORNEY/AGENT INFORMATION:
:   NAME: Marschang, Diane L.
:   REGISTRATION NUMBER: 35,600
:   REFERENCE/DOCKET NUMBER: P0755779P1
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: 415/225-5416
:     TELEFAX: 415/952-9881
:     TELETYPE: 910/371-7168
:   INFORMATION FOR SEQ ID NO: 15:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 10596 base pairs
:       TYPE: Nucleic Acid
:       STRANDEDNESS: Single
:       TOPOLOGY: Linear
:   US-08-087-783A-15

Query Match      2.9%; Score 67.2; DB 1; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggtggaaggaaggaagcccccgagaccacagagccagcccaaggtc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2287 GCGAGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGGCGCAGAGAGG 2346

QY 1016 ggaactgtggccagagatagcaaggtgg---ggcagggtggaaggggccccagc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2347 AGGGGACAGAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2406

QY 1073 caccgaagctggagcccccagagatgcggaacccttgaaagagagccagggtatagagcag 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2407 AGGAGCAGAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2466

QY 1133 ggggcccaggggaagatagccggagacccttaagccccaaagagagcaagaagaggaagc 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2467 GCGAGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2526

QY 1193 tggagcttgagcccgccgagagcagccgcccacagagcccccagaggtgctccacaggg 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2527 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2586

QY 1253 tggagaaatgcctctgaatttggaaggggtgtgcctcagccagggcagcctcaggaacgg 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2587 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2646

QY 1313 ggaaccagaaagtggcggtcaaggaaccttggggagggcaggtgcagccctgcgcgaacccc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2647 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2706

QY 1373 tggagaccaggggtggccgacaaggtgaaagagcggaggaaggttgataggggtgctgggg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2766

RESULT 11
US-08-194-088B-15
: Sequence 15, Application US/08194088B
: Patent No. 5580963
: GENERAL INFORMATION:
:   APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
:   TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
:   NUMBER OF SEQUENCES: 21
:   CORRESPONDENCE ADDRESSES:
:     ADDRESSEE: Genentech, Inc.
:     STREET: 460 Point San Bruno Blvd
:     CITY: South San Francisco
:     STATE: California
:     COUNTRY: USA
:     ZIP: 94080
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: patin (Genentech)
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/194,088B
:     FILING DATE: 09-FEB-1994
:     CLASSIFICATION: 530
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/884811
:       FILING DATE: 18-MAY-1992
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Gallegos, R. Thomas
:     REGISTRATION NUMBER: 32,692
:     REFERENCE/DOCKET NUMBER: 755D1
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 415/225-2614
:       TELEFAX: 415/952-9881
:       TELETYPE: 910/371-7168
:   INFORMATION FOR SEQ ID NO: 15:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 10596 bases
:       TYPE: nucleic acid
:       STRANDEDNESS: single
:       TOPOLOGY: Linear
:   US-08-194-088B-15

Query Match      2.9%; Score 67.2; DB 1; Length 10596;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 229; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 956 ggaaggtggaaggaaggaagcccccgagaccacagagcccaagagcccaaggtc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2287 GCGAGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2346

QY 1016 ggaactgtggccagagatagcaaggtgg---ggcagggtggaaggggccccagc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2347 AGGGGACAGAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2406

QY 1073 caccgaagctggagcccccagagatgcggaacccttgaaagagagccagggtatagagcag 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2407 AGGAGCAGAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2466

QY 1133 ggggcccaggggaagatagccggagacccttaagccccaaagagagcaagaagaggaagc 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2467 GCGAGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2526

QY 1193 tggagcttgagcccgccgagagcagccgcccacagagcccccagaggtgctccacaggg 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2527 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2586

QY 1253 tggagaaatgcctctgaatttggaaggggtgtgcctcagccagggcagcctcaggaacgg 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2587 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2646

QY 1313 ggaaccagaaagtggcggtcaaggaaccttggggagggcaggtgcagccctgcgcgaacccc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2647 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2706

QY 1373 tggagaccaggggtggccgacaaggtgaaagagcggaggaaggttgataggggtgctgggg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 AGGGGACAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGGCGCAGAGAGG 2766

RESULT 12
US-08-194-087-15
: Sequence 15, Application US/08194087
: Patent No. 5879910
: GENERAL INFORMATION:
:   APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
:   TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
:   NUMBER OF SEQUENCES: 21
:   CORRESPONDENCE ADDRESSES:
:     ADDRESSEE: Genentech, Inc.
:     STREET: 460 Point San Bruno Blvd
```


[illegible]

RESULT 14
US-07-945-

Sequence 1, Application US/07945283
 Patent No. 5352596
 GENERAL INFORMATION:
 APPLICANT: Cheung, Andrew K.
 APPLICANT: Wesley, Ronald D.
 TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
 TITLE OF INVENTION: Involving The Ep0 and LfT Genes
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtis P. Ribando
 STREET: 1815 No. 5352596th University Street
 CITY: Peoria
 STATE: IL
 COUNTRY: USA
 ZIP: 61604
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/945,283
 FILING DATE: 19920911
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Ribando, Curtis P
 REGISTRATION NUMBER: 27976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 309-685-4011 ext.513
 TELEFAX: 309-685-4128
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8438 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Pseudorabies virus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 622..6495
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(1099, "g")
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(1267, "t")
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(1381, "c")
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(1566, "c")
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(7010, "g")
 US-07-945-283-1

Query Match 2.68; Score 61.8; DB 1; Length 8438
 Best Local Similarity 67.48; Pred. No. 0.00033;

[illegible]

RESULT 15

```

US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

```

```

Query Match      2.6%   Score 60.2: DB 4:   length 3489;
Best Local Similarity 57.8%   Pred. No. 0.00053;
Matches 107; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY      2  agattcagcgccgcgttaattcttcgcgcagcatgacgacccgcagcactccact 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1028 AGGATGACGAGAGATGACGAGGAGATGACGAGAGAGATGACGAGGAGATGACGAGG 1087

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 18:53:59 ; Search time 17897.3 Seconds
(without alignments)
300.652 Million cell updates/sec

Title: US-09-431-843b-13
Perfect score: 1232
Sequence: 1 tagaatcagcgccgcctga.....attcattgacttgcgaataa 1232

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_dal: *
2: gb_dal2: *
3: gb_ov: *
4: gb_ov2: *
5: gb_pat: *
6: gb_ph: *
7: gb_p11: *
8: gb_p12: *
9: gb_p13: *
10: gb_p14: *
11: gb_p15: *
12: gb_p16: *
13: gb_p17: *
14: gb_p18: *
15: gb_p19: *
16: gb_p20: *
17: gb_p21: *
18: gb_p22: *
19: gb_p23: *
20: gb_p24: *
21: gb_p25: *
22: gb_p26: *
23: gb_p27: *
24: gb_p28: *
25: gb_p29: *
26: gb_p30: *
27: gb_p31: *
28: gb_p32: *
29: gb_p33: *
30: gb_p34: *
31: gb_p35: *
32: gb_p36: *
33: gb_p37: *
34: gb_p38: *
35: gb_p39: *
36: gb_p40: *
37: gb_p41: *
38: gb_p42: *
39: gb_p43: *
40: gb_p44: *
41: gb_p45: *
42: gb_p46: *
43: gb_p47: *

44: gb_htg9: *
45: gb_htg10: *
46: gb_htg11: *
47: gb_htg12: *
48: gb_htg13: *
49: gb_htg14: *
50: gb_htg15: *
51: gb_htg16: *
52: gb_htg17: *
53: gb_htg18: *
54: gb_htg19: *
55: gb_htg20: *
56: gb_htg21: *
57: gb_htg22: *
58: gb_htg23: *
59: gb_htg24: *
60: gb_htg25: *
61: gb_htg26: *
62: gb_htg27: *
63: gb_htg28: *
64: gb_htg29: *
65: gb_htg30: *
66: gb_htg31: *
67: gb_htg32: *
68: gb_htg33: *
69: gb_htg34: *
70: gb_htg35: *
71: gb_htg36: *
72: gb_htg37: *
73: gb_htg38: *
74: gb_htg39: *
75: gb_htg40: *
76: gb_htg41: *
77: gb_htg42: *
78: gb_htg43: *
79: gb_htg44: *
80: gb_htg45: *
81: gb_htg46: *
82: gb_htg47: *
83: gb_htg48: *
84: gb_htg49: *
85: gb_htg50: *
86: gb_htg51: *
87: gb_htg52: *
88: gb_htg53: *
89: gb_htg54: *
90: gb_htg55: *
91: gb_htg56: *
92: gb_htg57: *
93: gb_htg58: *
94: gb_htg59: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	1209	98.1	1232	11	AF172449	AF172449 Homo sapi
2	956	77.6	1676	11	AF172450	AF172450 Homo sapi
3	956	77.6	2423	11	AF172451	AF172451 Homo sapi
4	956	77.6	2483	11	AF172452	AF172452 Homo sapi
5	955.6	77.6	2363	11	AF172453	AF172453 Homo sapi
6	879.6	71.4	2423	11	AF109134	AF109134 Homo sapi
7	574	46.6	2208	12	AF156878	AF156878 Rattus no
8	325.4	26.4	160241	35	HS88567	AL035669 Human DNA
9	319.2	25.9	6670	11	AF112980	AF112980 Homo sapi
10	102	8.3	20083	91	HSCOL9A351	AF026801 Homo sapi
11	66.8	5.4	13680	73	HS2ULR	D10471 Herpes simp
12	66.8	5.4	154746	73	HSV2HG52	286099 Herpes simp

C	13	66.8	5.4	154746	73	HSV2HG52	Z86099 Herpes simp
C	14	66.4	5.4	236	91	HSRW62C3	X76589 H.sapiens D
C	15	65.4	5.3	214721	46	AC024957	MS004957 Mus muscu
C	16	65	5.3	15141	73	SHIPOIRE	M34651 Pseudocapre
C	17	64.8	5.3	124221	93	AC004825	AC004825 Homo sapi
C	18	64.8	5.3	161676	45	AC023461	AC023461 Homo sapi
C	19	64	5.2	173251	49	AC058798	AC058798 Homo sapi
C	20	63.6	5.2	74877	7	AB006705	AB006705 Arabidops
C	21	63.6	5.2	174011	68	AL355861	AL355861 Homo sapi
C	22	63.2	5.1	13095	91	HSFGRC4	Y13901 Homo sapien
C	23	62.8	5.1	1865	3	RABCBP	J05138 Rabbit calc
C	24	62.6	5.1	126055	71	HS73307	AL533999 Homo sapi
C	25	62.2	5.0	167976	47	AC026513	AC026513 Homo sapi
C	26	61.8	5.0	5123	72	HEPVYE	X15120 Pseudocapre
C	27	61.8	5.0	8438	73	SHLIT	M57505 Pseudocapre
C	28	61.8	5.0	68130	9	AC004961	AC004961 Homo sapi
C	29	61.6	5.0	205141	42	AC018679	AC018679 Homo sapi
C	30	61.2	5.0	460	12	MUSIRF3	M10296 Mouse DNA w
C	31	61	5.0	292668	71	HSICF13	AJ293930 Homo sapi
C	32	60.6	4.9	159984	41	AC065404	AC065404 Homo sapi
C	33	60	4.9	106664	12	AF130357	AF130357 Mus muscu
C	34	60	4.9	174717	42	AC018991	AC018991 Homo sapi
C	35	60	4.9	176364	37	AC009785	AC009785 Homo sapi
C	36	60	4.9	208980	49	AC053470	AC053470 Mus muscu
C	37	60	4.9	214194	52	AC073757	AC073757 Mus muscu
C	38	60	4.9	214574	42	AC019272	AC019272 Mus muscu
C	39	60	4.9	266875	52	AC073825	AC073825 Mus muscu
C	40	59.2	4.8	158336	67	AL139411	AL139411 Homo sapi
C	41	59.2	4.8	161175	52	AC073758	AC073758 Mus muscu
C	42	59.2	4.8	166074	45	AC023571	AC023571 Homo sapi
C	43	59	4.8	5286	12	RNDNABSP	X86100 R.noirveglicu
C	44	59	4.8	174533	52	AC073778	AC073778 Mus muscu
C	45	58.6	4.8	407	73	SHD737489	U27489 Suid herpes

ALIGNMENTS

	BASE COUNT	256 a	369 c	403 g	204 t	
	ORIGIN					
QY	1	tagaattcaagcgccgcctgaattctctagccgagcatgtgacgaaccccgactgcagctccac	60			
Db	1	TAGAATTAGGGGGCGCGCTGAATTCATGCCGAGCATGTGACGACGCCGACTG-CGACTCCAC	59			
QY	61	ctggagagagagacagagagatgcgagagaaacgagagagacagagactgcgagagacgagcga	120			
Db	60	CTGGAGAGAGAGACGAGAGATGCGGAGAGACGCCGAGAGACAGGAGACTGCGAGAGACGGCGA	119			
QY	121	gagcgacgagcgagagagacgcggaacgcagagagagacagagagatctcgagagagccgcg	180			
Db	120	gagcgacgagcgagagagacgcggaacgcagagagagacagagagatctcgagagagccgcg	179			
QY	181	gagcgacgagcgagcgactgcttcacgtccagagatcacagagagagatcgccagaacttcgagccac	240			

RESULT	1			
LOCUS	AF172449			
DEFINITION	AF172449	1232 bp	mRNA	PRI 19-APR-2000
ACCESSION	AF172449			
VERSION	AF172449.1	GI:7595298		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Emmalyotla, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 1232)			
TITLE	zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P. J. Cloning, sequencing, chromosomal location, and function of CDNAS encoding an oploid growth factor receptor (OGFR) in humans			
JOURNAL	Brain Res. 856 (1-2), 75-83 (2000)			
MEDLINE	20143540			
REFERENCE	2 (bases 1 to 1232)			
AUTHORS	zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P. J.			
TITLE	Direct Submission			
JOURNAL	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA			
FEATURES	Location/Qualifiers			
source	1..1232			

CDS
/cassette-type= placenta
34..1212
/note="Met-5}-enkephalin binding protein; OGFrr;
alternatively spliced"

QY	660	caacttcgcacataacagcatccccaagtcgcgtgtagtgtagcctgaagacattcca	719
Db	660	CAACTCCGCATCAACAGCATCTCTCAAGTCCGCCCTGTGAGCTAGACCTTCCA	719
QY	720	gagcgcaactggtccgctctctctctcgaggaagcgtgtgctgcgcggaagctgcgcggggt	779
Db	720	GAGCGCAACTGGTCCGCTTCTCTCTCTCGAGGAACCCCTGATGTGGCGGAGCTGCCGGGGGT	779
QY	780	gcgcgacagatgccctcgagactactcattgcttcgcgcgtgcgcgtccgcgcgaacacagccgcga	839

|||||
Db 780 GCGGAGAGTGCCTGAGACTACTCATGTTCCGCTGCGCTCCGACACACAGCCGCGCA 839
Oy 840 gctgtgtaacttcgcttgagagcaacttcggcccgctgcaagttcgttggggccca 899
Db 840 GCTGTGACACTTGGCTGGAGACACTCCGCGCCGCTCAAGTTGCTGGGGGCCCA 899
Oy 900 agacaagctgcagaggttcaagccagctctctgcagacacccgctcagaggtccagaa 959
Db 900 AGACAAGCTGCGAGAGTTCAACCCAGCTCTGCGCATCCGCTCGAGAGCTCCAGAA 959
Oy 960 ggtggaagaggaagacctgcagggagcagcagccagagccatcggaagcccaag 1019
Db 960 GGTGAGAGAGAGACCTGCGAGGGAGACGACCCGAGACCCATCGAGAACCCCAAG 1019
Oy 1020 ccccgagcccgagagactacaaagagatgagccaaagggggggggagagagccg 1079
Db 1020 CCGCAGCCCGGAGACCTCAAGAGGATGAGCCAGCCAAAGGGGAGGAGAGCCCG 1079
Oy 1080 ctgcctgctgtgtcttccaccagctctccctgcgcccctgtctgttaacagac 1139
Db 1080 CTGCTGCTGTGTCTTCCACCAAGCTCTCCCTGCGCCCTGTCTTGTATATGACC 1139
Oy 1140 ctctcgagcgggggcgagcgagcggtcttcttctgagtcgagtcgaagagagc 1199
Db 1140 CTCTGAGAGCGGGGCGGCGGAGGCGCTTCTTCTTATGATGATGCAAGCAAGC 1199
Oy 1200 ctcttcgaataatcatttacttctgaaaa 1232
Db 1200 CTTTCTGATAATTAATTCATTGACTTTCAGAAA 1232

RESULT 2
AF172450 1676 bp mRNA PRI 19-APR-2000
DEFINITION Homo sapiens opioid growth factor receptor mRNA, complete cds.
ACCESSION AF172450
VERSION AF172450.1 GI:7595300
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Cloning, sequencing, chromosomal location, and function of CDNA
encoding an opioid growth factor receptor (OGFr) in humans
JOURNAL Brain Res. 856 (1-2), 75-83 (2000)
MEDLINE 20143540
REFERENCE 2 (bases 1 to 1676)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Direct Submission (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA
FEATURES
Source
1. 1676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue_type="placenta"
34. 1419
/note="Met-5-enkephalin binding protein; OGFr;
alternatively spliced"
/codon_start=1
/product="opioid growth factor receptor"
/protein_id="AAF64403.1"
/db_xref="GI:7595301"
/translation="MDPPDCSTWEDDEDAEDAEDECEGEAGARADAGDEDE
SEEPRAAPSSFOSRMGSRNWRATRDRCRTNHYPDLYVEDNGDPNLSFYNEIR
FLPNGCFIEDILQWTDNYDILLEDNHSITQWLPRLREGVMAHAKPLTLREVYFKS
QEIORLVRAYELMIGFYGRLIEDRGITGVRAQNYQRFFONLMWRSHNNLRITRIK

SPCELSLEHFOAPLVREFLEETLVRELPGVRSALDYFMFAVCRHOROLYFANE
HFRRCFVWGPDKLRFKPSLSLPHLEDSRKVEEYSGSDPDHEHSTGRCGPB
SKGGKRVDEGPOPSVPEKPPDASSLPHLEDSRKVEEYSGSDPDHEHSTGRCGPB
EQPTEBPOBASVEXIALNLEECALISGSLRIGTQVVGQDPEASCPCKRWGLR
SCCLPFGSAS"
BASE COUNT 340 a 501 c 575 g 260 t
ORIGIN
Query Match 77.6%; Score 956; DB 11; Length 1676;
Best Local Similarity 99.3%; Pred. No. 9.7e-142;
Matches 981; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
Oy 1 tagaattcagcgccgcgcgaattctcagccagacatgagcagcccgactgncactcac 60
Db 1 TAGAATTACGCGCGCGCTGAAATTTAGCCGACATGACGACCCGACTG-CGACTTCAC 59
Oy 61 ctggaagagagagagagagatgcgagagacgcgagagacagagactgcagagacgca 120
Db 60 CTGGAGAGAGAGAGAGAGATGCCGAGAGCGCGAGAGACGAGACTGCGAGAGCGCGA 119
Oy 121 ggc 180
Db 120 GCGCCCGCGCGAGAGAGCGGAGCGGAGCGAGGAGACGAGAGTGGAGAGCCCGC 179
Oy 181 ggc 240
Db 180 GC 239
Oy 241 gagagacatgtgtgagatcgcgcacaactccgcgatctgtgtgagacagactgaatg 300
Db 240 GAGGACATGTTAGTATGCTCGGACCAACTATCCGATCTGGTGGAGACAGACTGCAATG 299
Oy 301 ggaacagccaaacctgagttctctacagaatgagatccgcgtctctgcgcaagcgttt 360
Db 300 GACACAGCCAAACCTGATTTCTACAGAAATGATCCGCTTCGCGCCCAAGCGCTGTT 359
Oy 361 catltagaacattctttagaacctgagcagacaatcactcctcttggagaaatcacatc 420
Db 360 CATTAGGACATCTTCAAGAACTGGACGACACACTATACCTCTTGAGGACAACTCATC 419
Oy 421 ctacatccagtgctgttctctctcgcgaaacagagtgagactgacatgcgaagccct 480
Db 420 CTACATCCAGTGGCTTTCCTCTCGTCGAGAACGAGATGAACTGCATGCCAAGCCCT 479
Oy 481 caagctcagagagtgagtgagtggtttaaagctccagagagatccagagcggtgttcg 540
Db 480 CACGCTCAGGAGGTGAGTGTAAAGCTCCAGAGATCCAGAGAGCGGCTGTCCG 539
Oy 541 ggcctcagagctcatagctgggcttctcagagatccgcgtgtagagaccagagcagggac 600
Db 540 GGCCTACAGCTCATGCTGGGCTTCTACGGGATCCGGGTGAGGACCCAGGACAGCGGAC 599
Oy 601 gttggcgagacagaaactccagaagcgctt-cagaactgaaactggcgacagcaaa 659
Db 600 GGTGGCGGAGACAGAACTACAGAAACGCTTCAGAACTGAACCTGGCCGACGACAA 659
Oy 660 caactccgcatcaacgcacatcctcaagtgcgcgtgtgagctgagcctcgacactcca 719
Db 660 CAACCTCCGATCAACACGATCTCAAGTCCGCGGTGAGCTGAGACCTCGAGACTTCCA 719
Oy 720 ggcgcacatgttcgctctctcctcggagagagagcgtgtgcgggagcgcgggggt 779
Db 720 GCGCGCACTGTGCGCTTCTCTGAGAGAACGCTGTGGGGGAGCTGCGGGGGGT 779
Oy 780 ggcgcagagtgccctgagactactcatgttcgcgcgtgctgcgcagacacagcgccgca 839
Db 780 GCGGAGAGTGCCTCGAGACTTATTTGCGCTGCGCTGCGGACACAGCGCGCGCA 839
Oy 840 gctgtgtaacttcgcttgagagcaacttcggcccgctgcaagttcgttggggccca 899
Db 840 GCTGTGACACTTGGCTGGAGACACTCCGCGCCGCTCAAGTTGCTGGGGGCCCA 899

QY	900	agagcagctcgaggaagttcaagcccaagctcttcgccgcatcgctcgagggtcccaagaa	959
Db	900	AGCAAGCTCGCGGAGGTTCAGGCGCAGCTCTTCGCCGATCCGCTCGAGGGCTCCAGGAA	959
QY	960	gctgaggaaggaagcctcgagggagac	987
Db	960	GGTGAGGAGGAAGGAAGCCCGGGGAC	987
RESULT	3		
AF172451			
LOCUS	AF172451	2423 bp	mRNA
DEFINITION	Homo sapiens	oploid growth factor receptor mRNA, complete cds.	PRI 19-APR-2000
ACCESSION	AF172451		
VERSION	AF172451.1	GI:7595302	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 2423)		
TITLE	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.		
JOURNAL	Cloning, sequencing, chromosomal location, and function of cDNAs encoding an oploid growth factor receptor (OGFr) in humans		
REFERENCE	Brain Res. 856 (1-2), 75-83 (2000)		
AUTHORS	2 (bases 1 to 2423)		
TITLE	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.		
JOURNAL	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA		
FEATURES	Location/Qualifiers		
SOURCE	1..2423		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="20"		
	/map="20q13.3"		
	/tissue_type="Placenta"		
	34..2067		
	/note="(Met-5)-enkephalin binding protein; OGFr; alternatively spliced"		
	/codon_start=1		
	/product="oploid growth factor receptor"		
	/protein_id="AA664404.1"		
	/db_xref="GI:7595303"		
	/translation="(MDPDCDSTWEDEEDADEDECEDEGAAGADADAGADEDEE SEEPRAARPSFOSRMTRSRMRATRDMCRGRHNYPDLYVEDGNDPINSFYNEIR FLPNGCFIEDILNMTDNYDILLEDNHSYIOMFLPLRPGVWMAKPLTIREVFKSS OEIOELRYARAEIYMGFYGLDEDRGTCTYVARAONYKRRPONLMWRSHNNRITRIIK SPCELSLEHFOAPLVRFPLETILVRELPGVROSALDIFMVAICRQROLVFAWEH HFRRCRFTVGWPQDKLRKFKPSLPHPLEGSRKVEEGSPEDPHKEASTOGRTPGPHS SKGGGRVDEGPOPSRVSVEPODAGPERSQDGAGGHEDEPRLSPKESKRKLISRR EOPRTGEPQASVEVERIALNLESCALSOGSLRTGTOEVGSDGEANVOPCROPGLGAR VADYVRRKRRKDEGAGDAAVAGSAGOTIALAGSPAGGHPKAGHSNVSVEDEEGTTS GAKPRTGSGSPETSPGSPAGPAGDEPAESPESETGSPAGPAGACGPAPESBETGLAPR AGPAGDEPAETSPETSPGSPAGPTRDEPAESPESTPGPRPAGPAGDEPAESBETGPB RPAEPAGDEPESSETPGSPAGPAPTRDEPAKAEAEALDPAEVSSAKSGKP"		
BASE COUNT	503 a	754 c	843 g
ORIGIN	323 t		
Query Match	77.6%	Score 956:	DB 11:
Best Local Similarity	99.3%	Pred. No. 8.5e-142;	Length 2423;
Matches	981;	Conservative 0;	Mismatches 5;
		Indels 2;	Gaps 2;
QY	1	tagaattcagcgccgcctcgaattcagcccgagcatgacgaccccgactgncactcac	60
Db	1	TAGAATTACGCGCGCGCTGAATTCTACGCCGAGCAICAGACACC CGAATG-CGATCTCAC	59
QY	61	ctggaggaggaagacgagggatcgaggagacgcggaggaagcagaagactcgcagagcgca	120
Db	60	CTGGGAGGAGGACGAGGAGATGCGGAGACGCGGAGGACGACGAGCACTCCGAGGACGGCGGA	119

[illegible]

RESULT	4
AF172453	
LOCUS	AF172453
DEFINITION	Homo sapiens oploid growth factor receptor mRNA, complete cds.
ACCESSION	AF172453
VERSION	AF172453.1
KEYWORDS	GI:7595306
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2483)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Cloning, sequencing, chromosomal location, and function of cDNAs encoding an opioid growth factor receptor (OGFR) in humans
JOURNAL Brain Res. 856 (1-2), 75-83 (2000)
MEDLINE 20143540
REFERENCE 2 (bases 1 to 2483)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA
FEATURES
source 1. .2483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue_type="Placenta"
34. .2127
/note="[Met-5]-enkephalin binding protein: OGFR;
alternatively spliced"
/codon_start=1
/product="opioid growth factor receptor"
/protein_id="AAF64406.1"
/db_xref="GI:7595307"
/translation="MDPDCSTWEDDEDEDAEDAEDECEGEAAAGADADGDEDE
SEPRARPSFOSRMGSRWRATRDRCRTRHNPDLVEDDCNDTNSFTRKEIR
FLPNCFTLEDLWMTDNYDLEDNHSTIOWLPREGVNMAKPLTRLEVEFKSS
QEIQRRLRAYELMGEFYGRLEDRGTGVAQNYOKRPNLMRSHNNLRITRLK
SPCLSLFHEQAPLVRFLLETLVARELQYROSALDYFMAVRCRHRQLVHAME
HFRRCRVMPQKLRFRKPSLPHLEGSRYKEESPGDPDHEASTOGTCPEH
SKGGRVNDEGSPRSVERODAGPERSGSGDAGHGDEPRPLSPKSKRKLISLR
EQPTGEPGQSAEYERKALNLEGCALSGLSRGTGTVGGQDDEBAAPQPLGAK
VADVRRKRYKDEGDSAAVAGSAGQTLAAGSPABSGHRAKSHSENGVEDTEGRT
GPKETGSPSPETBPSPAGADGPAPATPSETBPSPAGPTREBPASBPETGPR
AGPADPEAPESPETGPRPAGPADPEPAKIPSETGSPAGPTREBPASBPETGPR
RPAGPADPEAPESPETGPRPAGPADPEAPESPETGPRPAGPTREBPAGAAAE
LQDAVESSASNGRP"
BASE COUNT 518 a 778 c 861 g 326 t
ORIGIN

Query Match 77.6%; Score 956; DB 11; Length 2483;
Best Local Similarity 99.3%; Pred. No. 8.4e-142;
Matches 981; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 1 tagaatcaagcgagcgtgaattctagccgagcatgagaccccgactgncgactccac 60
DB 1 TAGAATCAGCGCGCGCTGAATTCTAGCCGACATGAGACGCCCGACTG-CGACTCCAC 59

QY 61 ctggagagagagcgagagatgctgagagagcgagagagagagagagagagcgaga 120
DB 1 CTGAGATCAGCGCGCGCTGAATTCTAGCCGACATGAGACGCCCGACTG-CGACTCCAC 59

QY 60 CTGGAGAGAGAGCGAGAGATGCGAGAGACCGAGAGCGAGAGCTGCGAGAGCGGCA 119

QY 121 ggc 180
DB 120 GCGCGCGCGCGCGAGAGCGCGAGCGGAGCGAGAGCGAGAGCGAGAGCGCGCG 179

QY 181 ggc 240
DB 180 GCGCGCGCGCGCGAGAGCGCGAGCGGAGCGGAGCGGAGCGGAGCGCGCG 239

QY 241 gag 300
DB 240 GAGGAG 299

QY 301 gag 360
DB 300 GAGGAG 359

QY 361 cattag 420
DB 361 CATTAG 419

Db 360 CATTAG 419
QY 421 ctactcag 480
DB 420 CTACTCAG 479

QY 481 cagcgtcag 540
DB 480 CAGCGTCAG 539

QY 541 ggcctcag 600
DB 540 GGCCTCAG 599

QY 601 ggtggccgag 659
DB 600 GGTGGCCGAG 659

QY 660 caactcgcag 719
DB 660 CAACTCGCAG 719

QY 720 ggc 779
DB 720 GCGCGCGCGCGCGAGAGCGCGAG 779

QY 780 ggc 839
DB 780 GCGCGCGCGCGCGAGAGCGCGAG 839

QY 840 gctgtgag 899
DB 840 GCTGTGAG 899

QY 900 agacag 959
DB 900 AGACAG 959

QY 960 ggtgag 987
DB 960 GGTGAG 987

RESULT 5
AF172452 2363 bp mRNA PRI 19-APR-2000
LOCUS Homo sapiens opioid growth factor receptor mRNA, complete cds.
DEFINITION AF172452
ACCESSION AF172452
VERSION AF172452.1 GI:7595304
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2363)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Cloning, sequencing, chromosomal location, and function of cDNAs encoding an opioid growth factor receptor (OGFR) in humans
JOURNAL Brain Res. 856 (1-2), 75-83 (2000)
MEDLINE 20143540
REFERENCE 2 (bases 1 to 2363)
AUTHORS Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA
FEATURES
source 1. .2363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue_type="Placenta"

OY	247	cgtgtagtgatcatcggccaactatcccgtaattcttgtgaacagaaactcycaatyyggaac	306
Db	333	CATGCAAAAGTATACGGCACACTCCCGGATTGTGACAGATCTAAGACTCATAATGGTGAT	392
OY	307	gccaaaaccttgaagtcttcacaagaatvagaaccgcctcccgtcccaaacygttttcattga	366
Db	393	GTCACAACCTGAGGTTCTTACAAAATGAGATCTGCTTCAGGCCAAATGGGGCTCTCATCGA	452
OY	367	ggagcatcttcagaacttgaacggaacatavagactcccttgaaggacaatcacctccaat	426
Db	453	GGACATTCCTTCANAACTGGAAAAAGACAACTATACCCTCTGGGAAGAATCACTCCTCAT	512
OY	427	ccagttgcgttttccccctctcggaaacacaggagttgaacttggcatgtccaaagcccccacagt	486
Db	513	CCAGTGGCTGTTTCCCTCTCGGGGAAACAGAGATGAATCTGGACGCCAACGCCCTTCACCT	572
OY	487	cagggaagtcagaagtygtttaaaagctcccaaggagatlccaaagagcgycttgtccagycta	546
Db	573	GAAAGGAGTGTGAGGCATTAAAAAGCTCCAAGGAAATCGAAGAGCGTCTTCGCGGGCTTA	632
OY	547	cgagaccatgcctcggagctcttaaggatccogcttggaggaccggagacagggcaagtyggg	606
Db	633	TGAGCTCATAGCTGGGCTTTATGGGTTTTACCTTAGGATGAGGACCGGGGCAAGGGTGTATG	692
OY	607	ccgagcacagaactcccaagaagcgtt--cagaacctgaactctygcgcagcccaacaacct	665
Db	693	CCGTGCACAGAACTTCCAGCCCCGCTTCACATCTGAACAGCCACAGCCACAAACACT	752
OY	666	ccgcatcacagaactccatccaaatgcgcgtlfglyagatcgtctgaagcaattccaggycgc	725
Db	753	GCGTATTACACCACTCTCMACTGCTGGGAGTGAGTGGGCTTAAACAACTACACGGGCACC	812
OY	726	actgtgtccgcttcttctcttggaggaaacgctgtgcgcgcggagactcccggygtgcgca	785
Db	813	CTGTGGCCSCTTCTCTGAGAGAACCTTGTAAGACAAACTGGCCAGCGTGCGCCA	872
OY	786	gggtgcctctyggaacttatcattgttcocgtgccttgcgcagacaacgcgcgcgaactggt	845
Db	873	GAGTCCCCTGGACNTCTTCTGTCGCTGCGTCCGTCGGCACACGCGCGGGAGCTTGT	932
OY	846	gaacttgcctcttgggaacttccgcgcgcgcgcgttgaagtctctcggggggcccaagaaca	905
Db	933	GTACTTTGCGTGGGAGCACTTCAGACCCTCGCCGAGAGATTGTCTGSGGGGCCCTGTACAA	992
OY	906	gctcgcggaggttcaaagccagctctctgcgcgatccgcctcgaaggtctccagaagtttga	965
Db	993	GCTGGGGAGATTCAAAGCCCCAGACCATACCCCACTGACGGGAGCCAGGCGAGCGACAGA	1052
OY	966	ggaggaaggaacctgcaggggagc	987
Db	1053	TAAAGTATGAGGCTCCAGGGAC	1074
RESULT	8		
LOCUS	HS885L7		
DEFINITION	Human DNA sequence from Clone RP5-885L7 on chromosome	PRI	09-JUL-2000
ACCESSION	AL035669		
VERSION	AL035669.43		
KEYWORDS	HIT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheraia; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	I (bases 1 to 160241)		
JOURNAL	Direct Submission Submitted (09-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Jul 8, 2000 this sequence version replaced gi:8919619. During sequence assembly data is compared from overlapping clones		
COMMENT			

FEATURES

Source

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Embl, EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP, information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone comp15 of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

RP5-88517 is from the library RPCI-5 constructed at the Rosewell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-88517 The true left end of clone RP4-563E14 is at 140192 in this sequence. The true right end of clone RP11-93B14 is at 44512 in this sequence.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at <http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20> RP5-885L7 is from the library RPc1-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR>

This sequence is the entire insert of clone RP5-885L7 The true left end of clone RP4-563B14 is at 140192 in this sequence. The true right end of clone RP11-93B14 is at 44512 in this sequence.

```

TURS
source
    1..160241
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /chromosome="20"
       /map="q13.2-13.33"
       /clone="RP5-885L7"
       /clone_11b="RPc1-5"
    2..64
       /note="LIMa9 repeat: matches 6210..6254 of consensus"
    65..353
       /note="AluI repeat: matches 1..287 of consensus"
    354..397
       /note="LIMa9 repeat: matches 6254..6302 of consensus"
    4198..5157
       /note="12 copies 80 mer 61% conserved"
       /note="complement(4223..5193)"
       /note="random repeat: region contains two forced joins and
       single clone. Assembly consistent with restriction
       digest."
    4229..5114
       /note="443 copies 2 mer cc 59% conserved"
    4231..5166
       /note="234 copies 4 mer tccc 60% conserved"
    4240..5151
       /note="19 copies 48 mer 61% conserved"
    4248..5159
       /note="38 copies 24 mer 61% conserved"
    4264..5163
       /note="15 copies 60 mer 61% conserved"
    4286..5152
       /note="17 copies 51 mer 59% conserved"
    4308..5171
       /note="27 copies 32 mer 61% conserved"
    4317..5156
       /note="21 copies 40 mer 62% conserved"
    4344..5142
       /note="17 copies 47 mer 59% conserved"
    4348..5145
       /note="42 copies 19 mer 60% conserved"
    4431..5144
       /note="21 copies 34 mer 60% conserved"
    4456..5148
       /note="21 copies 33 mer 60% conserved"
    4486..5151
       /note="9 copies 74 mer 69% conserved"
```

```
repeat_region 4530..5087
/note="18 copies 31 mer 71% conserved"
repeat_region 4533..4822
/note="5 copies 58 mer 77% conserved"
repeat_region 4720..5137
/note="11 copies 38 mer 61% conserved"
repeat_region 4833..5064
/note="4 copies 58 mer 79% conserved"
repeat_region 5051..5164
/note="19 copies 6 mer tctctc 63% conserved"
repeat_region 5934..6864
/note="LIM4 repeat: matches 4509..5488 of consensus"
repeat_region 6974..7049
/note="4 copies 19 mer 80% conserved"
repeat_region 7224..7451
/note="6 copies 38 mer 98% conserved"
repeat_region 7892..8074
/note="MER68A repeat: matches 68..210 of consensus"
repeat_region 8674..8966
/note="AluX repeat: matches 5..296 of consensus"
repeat_region 10876..10962
/note="MIR repeat: matches 86..162 of consensus"
repeat_region 11974..12025
/note="13 copies 4 mer ctca 73% conserved"
repeat_region 14394..14478
/note="5 copies 17 mer 77% conserved"
repeat_region 15824..15891
/note="17 copies 4 mer tccc 67% conserved"
repeat_region 16266..16577
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region 16578..16866
/note="17 copies 17 mer 65% conserved"
repeat_region 16581..16844
/note="11 copies 24 mer 68% conserved"
repeat_region 16582..16891
/note="10 copies 31 mer 66% conserved"
repeat_region 16583..16879
/note="9 copies 33 mer 66% conserved"
repeat_region 16585..16840
/note="16 copies 16 mer 66% conserved"
repeat_region 16586..16875
/note="5 copies 58 mer 72% conserved"
repeat_region 16590..16874
/note="15 copies 19 mer 64% conserved"
repeat_region 16593..16872
/note="7 copies 40 mer 67% conserved"
repeat_region 16598..16837
/note="5 copies 48 mer 73% conserved"
repeat_region 16603..16842
/note="3 copies 80 mer 77% conserved"
repeat_region 16638..16801
/note="44 copies 6 mer aaagaa 62% conserved"
repeat_region 16664..16885
/note="3 copies 74 mer 78% conserved"
repeat_region 16670..16804
/note="5 copies 47 mer 77% conserved"
repeat_region 16848..16899
/note="13 copies 4 mer gaag 100% conserved"
repeat_region 17287..17411
/note="MERSA repeat: matches 35..165 of consensus"
repeat_region 17447..17486
/note="10 copies 4 mer ccat 92% conserved"
repeat_region 18303..18502
/note="5 copies 40 mer 66% conserved"
repeat_region 18307..18498
/note="4 copies 48 mer 67% conserved"
repeat_region 18317..18496
/note="3 copies 60 mer 74% conserved"
repeat_region 18319..18486
/note="7 copies 24 mer 66% conserved"
repeat_region 18320..18495
/note="44 copies 4 mer aacac 64% conserved"
repeat_region 18321..18496
```

```
/note="11 copies 16 mer 64% conserved"
18325..18486
/note="9 copies 18 mer 66% conserved"
repeat_region 18327..18486
/note="8 copies 20 mer 66% conserved"
repeat_region 18497..18546
/note="25 copies 2 mer ac 72% conserved"
repeat_region 19329..19433
/note="LTR16C repeat: matches 286..387 of consensus"
repeat_region 19455..19590
/note="4 copies 34 mer 88% conserved"
repeat_region 19457..19604
/note="2 copies 74 mer 81% conserved"
repeat_region 19546..19625
/note="4 copies 20 mer 76% conserved"
repeat_region 19839..19934
/note="2 copies 58 mer 87% conserved"
repeat_region 20020..20103
/note="MIR repeat: matches 51..137 of consensus"
repeat_region 20704..20805
/note="3 copies 34 mer 90% conserved"
repeat_region 20710..20811
/note="2 copies 51 mer 90% conserved"
repeat_region 20722..20806
/note="5 copies 17 mer 89% conserved"
repeat_region 21349..21410
/note="2 copies 31 mer 93% conserved"
repeat_region 21584..21654
/note="LTR16A repeat: matches 88..158 of consensus"
repeat_region 21779..21852
/note="Alu repeat: matches 238..311 of consensus"
repeat_region 24787..24847
/note="MER91B repeat: matches 1..63 of consensus"
repeat_region 24855..25064
/note="LTR1D repeat: matches 107..348 of consensus"
repeat_region 25853..26065
```

```
Query Match 26.4% Score 325.4; DB 35; Length 160241;
Best Local Similarity 96.8%; Pred. No. 4.3e-43;
Matches 332; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 645 ctggcgacagcacaacacccatccatcacagcatcccaagtcgctgtgagctgag 704
|||||
Db 61404 CAGGCGCAGCCCAACAACCTCCGATCACACGATCTCAAGTCGCTGGTAGCTGGG 61463
|||||
QY 705 cctcgagcacttcagggcgccactggtccgtcttcctctgagagagacgctgtgctg 764
|||||
Db 61464 CCTCGAGCACTTCAGGCGCGCGTGCCTCTCTCTGTGAGAGAGAGCGTGTGGCGG 61523
|||||
QY 765 ggaagctgcgggggtgaggagatgcccctgactactcatgttcgcgtgctgcg 824
|||||
Db 61524 GGAGCTGCCGGGGGGGCGCGGAGATGCCCTGTGACTTCACTGTTCGCGCGCTCGCG 61583
|||||
QY 825 aacacagcgccgacgctgtgtacattgcctggaggagacttcgggcccgcgtcaagt 884
|||||
Db 61584 ACACCAAGCGCCGCCAGCTGTGTCACTTCCTGGAGACACTTCGGCCCTCAAGTT 61443
|||||
QY 885 cgtctggggccccaagaacagctgtcggaagttcaagccagctctctgcgcacatcgct 944
|||||
Db 61644 CGTCTGGGGGCCCAAGACAGCTGTGGAGGTCAAGGCCAGCTCTGTGCCCATCGCT 61703
|||||
QY 945 cgaagctcccaagaaagtgtgagaggaagacgtcgcaggagac 987
|||||
Db 61704 CGAGGCTCCAGGAAGGTGAGAGGAAGGCCCGGGGAC 61746
|||||
```

```
RESULT 9
AF112980 6670 bp DNA PRI 10-JAN-1999
LOCUS AF112980
DEFINITION Homo sapiens 7-60 gene, complete cds.
ACCESSION AF112980
VERSION AF112980.1 GI:4139271
KEYWORDS
```

```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       1 (bases 1 to 6670)
JOURNAL     Takanosu,M., Liu,J. and Wayne,R.
REFERENCE   Genomic Structure of Human Gene 7-60
AUTHORS     Unpublished
TITLE       2 (bases 1 to 6670)
JOURNAL     Takanosu,M., Liu,J. and Wayne,R.
REFERENCE   Direct Submission
TITLE       Submitted (10-DEC-1998) Cell Biology, University of Alabama at
JOURNAL     Birmingham, VH 302, Room 605, Birmingham, Alabama 35294, USA
FEATURES
  source
    1..6670
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="20"
    /map="20q13.3"
    join(<206..268,906..984,2235..2313,3147..3213,4125..4274,
4891..>6309)
    /product="7-60"
    join(206..268,906..984,2235..2313,3147..3213,4125..4274,
4891..6309)
    /codon_start=1
    /product="7-60"
    /protein_id="AAD03745.1"
    /db_xref="GI:4139272"
    /translation="MTGSRNWRATYDMCRHYRNPDLVERDNGDTPNLSPYRNEIR
LPNGCFIEDILQNTWDNDYDLLEDNHSYIOWLPLEPEVNMWAKPLTREVFEVSQ
EIQERLVRAVAAHAGILNDPAGPRHGHGSPSTLEPEALPEPLRSHNLRTTRIKS
LGEIQLHEHQADLVRFLESLVRELPGVRSALDYFMFAVGCRRHOROLVHFAMEH
FPRCKFTVWQPODKLRFKFSPLPHLGSRRVEEGRSGPDDEHSAVSGCTGCPHS
KGGRRDEGPOPSVPEODAGPLERSOGDEAGHEDDEPEPLSPKESKKLELRRE
OPPRRGPSSASEVEKIALINLEGCALSOGSLGTQVGGDDPGAVQPCQPLGARR
AKVRRKRVNDEGACDAAVAGGAQTALAGSPAPSGHPAGHSENGVEDTGGRTG
PREGTPGSPSTPSPAGPAGDEPAESPSTPSPAGPAGDEPAESPSTPSPRPA
GPAGDEPAKTPSETPGSPAGPTRDEPAESPSTPSPAGPAGDEPAESPSTPSPRPA
PAGPAGDEPAESPSTPSPAGPTRDEPAESPSTPSPAGPAGDEPAESPSTPSPRPA"

BASE COUNT      1367 a      1998 c      2169 g      1136 t
ORIGIN
Query Match      25.9%; Score 319.2; DB 11; Length 6670;
Best Local Similarity 96.2%; Pred. No. 1.3e-41;
Matches 327; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 648 ggcgcagccacacacactcgcgcatcacacgcatctcgaagtgcggtgagctgagcct 707
      |||||||
DB 4890 ggcgcagccacacacactcgcgcatcacacgcatctcgaagtgcggtgagcct 707
      |||||||
QY 708 cgaagcacttcagagcgccacgtgctcgtcttcctctgagagagagcgctgagcgagga 767
      |||||||
DB 4950 cgaagcacttcagagcgccacgtgctcgtcttcctctgagagagagcgctgagcgagga 5009
      |||||||
QY 768 gctgcgcgaggggtgcgagagagtgccctggaactcattcattcgctgcgtgcgctgcgaca 827
      |||||||
DB 5010 gctgcgcgaggggtgcgagagagtgccctggaactcattcattcgctgcgtgcgctgcgaca 5069
      |||||||
QY 828 ccaagcgcccgccagctggtgacacttcgctgagagacattccgagcccgctgaagttcgt 887
      |||||||
DB 5070 ccaagcgcccgccagctggtgacacttcgctgagagacattccgagcccgctgaagttcgt 5129
      |||||||
QY 888 ctggggggcccaagaagaagctgcgagagtttcaagcccgactctctgcgcgcatccgctcga 947
      |||||||
DB 5130 ctggggggcccaagaagaagctgcgagagtttcaagcccgactctctgcgcgcatccgctcga 5189
      |||||||
QY 948 ggggtccaggaaggtgagagagagagagagagagagagagagagagagagagagagagac 987
      |||||||
DB 5190 ggggtccaggaaggtgagagagagagagagagagagagagagagagagagagagagagac 5229
      |||||||

RESULT 10

```

```

HSCOL9A3S1
LOCUS      HSCOL9A3S1 20083 bp DNA
DEFINITION Homo sapiens alpha-3 type IX collagen (COL9A3) gene, promoter
            region, and exons 1-26.
ACCESSION AF026801
VERSION    AF026801.1 GI:5712084
KEYWORDS   1 of 2
SEGMENT    human.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       1 (bases 1 to 20083)
JOURNAL     Paassilta,P., Pihlajamäki,T., Annunen,S., Brewton,R.G., Wood,B.M.,
MEDLINE     Johnson,C.C., Liu,J., Gong,Y., Worman,M.L., Prockop,D.J., Wayne,R.
REFERENCE   and Ala-Kotko,L.
AUTHORS     Complete sequence of the 23-kilobase human COL9A3 gene. Detection
            of Gly-X-Y triplet deletions that represent neutral variants
            J. Biol. Chem. 274 (32), 22469-22475 (1999)
            99357778
            2 (bases 1 to 20083)
            Paassilta,P., Annunen,S., Brewton,R.G., Wood,B.M., Johnson,C.C.,
            Liu,J., Gong,Y., Worman,M.L., Prockop,D.J., Wayne,R. and
            Ala-Kotko,L.
            Direct Submission
            Submitted (25-SEP-1997) Department of Medical Biochemistry,
            University of Oulu, Kajaanintie 52 A, Oulu 90220, Finland
            Location/Qualifiers
            1..20083
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="COL9A3 gene from intron 26 to 3'UTR"
            1..3188
            /gene="COL9A3"
            complement(1838..2099)
            /rpt_family="Alu"
            <3189..3266
            /gene="COL9A3"
            /number=1
            3267..3679
            /gene="COL9A3"
            /number=1
            3680..3748
            /gene="COL9A3"
            /number=2
            3749..4624
            /gene="COL9A3"
            /number=2
            4625..4660
            /gene="COL9A3"
            /number=3
            4661..5321
            /gene="COL9A3"
            /number=3
            5322..5393
            /gene="COL9A3"
            /number=4
            5394..6024
            /gene="COL9A3"
            /number=4
            6025..6078
            /gene="COL9A3"
            /number=5
            6079..7272
            /gene="COL9A3"
            /number=5
            7273..7308
            /gene="COL9A3"
            /number=6
            7309..7596
            /gene="COL9A3"
            /number=6
            7597..7620

```

```
/gene="COL9A3"
/number=7
7621..7846
/gene="COL9A3"
/number=7
7847..7900
/gene="COL9A3"
/number=8
7901..8197
/gene="COL9A3"
/number=8
8198..8251
/gene="COL9A3"
/number=9
8252..8676
/gene="COL9A3"
/number=9
8677..8718
/gene="COL9A3"
/number=10
8719..10535
/gene="COL9A3"
/number=10
9151..9426
/rpt_family="Alu"
10536..10592
/gene="COL9A3"
/number=11
10593..11059
/gene="COL9A3"
/number=11
11060..11113
/gene="COL9A3"
/number=12
11114..11810
/gene="COL9A3"
/number=12
11811..11864
/gene="COL9A3"
/number=13
11865..12197
/gene="COL9A3"
/number=13
12198..12251
/gene="COL9A3"
/number=14
12252..12760
/gene="COL9A3"
/number=14
12761..12814
/gene="COL9A3"
/number=15
12815..13234
/gene="COL9A3"
/number=15
13235..13288
/gene="COL9A3"
/number=16
13289..13914
/gene="COL9A3"
/number=16
13915..13968
/gene="COL9A3"
/number=17
13969..14755
/gene="COL9A3"
/number=17
14756..14809
/gene="COL9A3"
/number=18
14810..14914
/gene="COL9A3"
/number=18

exon      14915..14968
           /gene="COL9A3"
           /number=19
           14969..15445
           /gene="COL9A3"
           /number=19
           15446..15490
           /gene="COL9A3"
           /number=20
           15491..15603
           /gene="COL9A3"
           /number=20
           15604..15657
           /gene="COL9A3"
           /number=21
           15658..15741
           /gene="COL9A3"
           /number=21
           15742..15795
           /gene="COL9A3"
           /number=22
           15796..16334
           /gene="COL9A3"
           /number=22
           16335..16388
           /gene="COL9A3"
           /number=23
           16389..16490
           /gene="COL9A3"
           /number=23
           16491..16562
           /gene="COL9A3"
           /number=24
           16563..17983
           /gene="COL9A3"
           /number=24
           17984..18019
           /gene="COL9A3"
           /number=25
           18020..18899
           /gene="COL9A3"
           /number=25
           18900..18944
           /gene="COL9A3"
           /number=26
           18945..>20083
           /gene="COL9A3"
           /number=26
           complement(19292..20080)
           /rpt_family="Alu"

repeat_region
BASE COUNT      3285 a      6190 c      6614 g      3994 t
ORIGIN

Query Match      8.3%: Score 102; DB 91; Length 20083;
Best Local Similarity 95.1%: Pred. NO. 1.4e-07;
Matches 116; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1105 gctctccctgcgccctgtctttgtatcgaacctcttgagcgggagcgagca 1164
    |||
DB 1 GCTCTCCCTGCGGCCCTGTCTTGTATAATGACCTTCTGAGTGGGGGGCGGCA 60

QY 1165 gggcttgcttcttagtcgtgcgaagcaaggcctttctgaataatcattgact 1224
    |||
DB 61 GGGC-TGCTTTTCTAGTGTATACCAAGCAAGCCTTCTGATAAATCATTTGACT 119

QY 1225 tt 1226
    ||
DB 120 TT 121

RESULT 11
HS2ULIR
```

LOCUS	HS2ULIR	13680 bp	DNA	VRL	14-APR-2000
DEFINITION	Herpes simplex virus type 2 genomic DNA for 0.74-0.84 region.				
ACCESSION	D10471 D01128				
KEYWORDS	ICP0; ICP27; ICP34.5; RL1; RL2; Syn-associated membrane glycoprotein; UL53; UL54; UL55; Vmw118; Vmw63; long repeat region; long unique region; neurovirulence factor; transcriptional regulator.				
SOURCE	Herpes simplex virus type 2 (isolate:HG52) DNA, clone:1f, p and g.				
ORGANISM	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Simplexvirus.				
REFERENCE	1 (sites)				
AUTHORS	Davison,A.J. and Wiltke,N.M.				
TITLE	Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2				
JOURNAL	J. Gen. Virol. 55 (Pt 2), 315-331 (1981)				
REFERENCE	2 (sites)				
AUTHORS	Whitton,J.L., Rixon,F.J., Easton,A.J. and Clements,J.B.				
TITLE	Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is conserved: conserved sequences around the 5' and 3' termini correspond to transcription regulatory signals				
JOURNAL	Nucleic Acids Res. 11 (18), 6271-6287 (1983)				
REFERENCE	3 (bases 1 to 13680)				
AUTHORS	McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.				
TITLE	Comparative sequence analysis of the long repeat regions and adjoining parts of the long unique regions in the genomes of herpes simplex viruses types 1 and 2				
JOURNAL	J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)				
FEATURES	92113549				
Source	Location/Qualifiers				
	1..13680				
	/organism="human herpesvirus 2"				
	/isolate="HG52"				
	/db_xref="taxon:10310"				
	/clone="f, p and g"				
	<2..412				
	/note="UL53 ORF"				
	/codon_start=1				
	/product="Syn-associated membrane glycoprotein"				
	/protein_id="BAA01268.1"				
	/db_xref="gi:221787"				
	/translation="DPVFLHRAIGIYVCCCELLRVALGLIVGALLSRGCAIT				
	HPLFLITTWCFVSIATLLELIFILIRKSAKNAEPAPRKRSGWSGVCGCCSILL				
	SGIAVRLCYIAVAVGAVLALRYEDELQRLDL"				
	460..465				
	/note="for UL53"				
	802..806				
	/note="for UL54"				
	958..2496				
	/note="UL54 ORF"				
	/codon_start=1				
	/product="Vmw63(ICP27)"				
	/protein_id="BAA01269.1"				
	/db_xref="gi:221788"				
	/translation="MATDIDMLDGLDLSSELEDALEDEEGRDDPESDGGEC				
	SSSDMEDPCGDGAELIDAIIPKGPAPEDGTPASTPPAARAGADDPAPT				
	GVMSRLGTRRSASPREHKGVARIOPTSTKAPRGRGRGRGRGADSP				
	KPRRVSNHANGGRHAPASARDGPGATHEARGGGLDVGSGPRGRAPPL				
	MALSLTPHADGRAPVPERKASADTIDPAVRAVLSISESAVERISSEGRALM				
	ODPGMGPEFPAASPMAPVLATQAGDPAETRRVSWETLVAHGSLVTRFANRAAS				
	TAKAMRCVLAQENLIELASADETLAWCKMCIHNPLRPDPITGTAAVLENLT				
	RLRPLDCLYLRAGLGGDLCSRRRLSDIDIASFVLVILARLANRVERGVSDYT				
	TVGVAGETMHFYIPGACMGLILELIDTHRECCSSRVCELTASHTIAPLVHGYEYC				
	NSLF"				
	2502..2507				
	/note="for UL54"				
	2711..3271				
	/codon_start=1				
	/product="UL55 ORF"				
polyA_signal					
TATA_signal					
CDS					
misc_feature					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					
misc_feature					
misc_feature					
polyA_signal					
CDS					

Query Match

Best Local Similarity 5.4%; Score 66.8; DB 73; Length 13680;

Matches 101; Conservative 0; Mismatches 59; Indels 0; Gaps 0.

TATA_signal /product="neovirulence factor(ICP34.5)"
complement(13338..13341)
misc_feature /note="for RL1"
13365..13618
misc_feature /note="terminally redundant element(a sequence)"
13619

BASE COUNT 1874 a 5124 c 4789 g 1893 t

ORIGIN 0.74-0.84 genome % units.

exon /number=2
/product="neovirulence factor(ICP34.5)"
complement(12685..>13179)

misc_feature /number=1
/product="neovirulence factor(ICP34.5)"
complement(13338..13341)
misc_feature /note="for RL1"

RESULT 12

HSV2HG52

LOCUS HSV2HG52 154746 bp DNA VRL 12-DEC-1999

DEFINITION Herpes simplex virus type 2 (strain HG52), complete genome.

ACCESSION Z86099

VERSION Z86099.2 GI:6572414

KEYWORDS capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyuridine triphosphatase; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein F; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; Rf1 gene; Rf2 gene; Rf1 gene; tegument protein; thymidine kinase; UL1 gene; UL10 gene; UL11 gene; UL12 gene; UL13 gene; UL14 gene; UL15 gene; UL16 gene; UL17 gene; UL18 gene; UL19 gene; UL20 gene; UL21 gene; UL22 gene; UL23 gene; UL24 gene; UL25 gene; UL26 gene; UL26.5 gene; UL27 gene; UL28 gene; UL29 gene; UL30 gene; UL31 gene; UL32 gene; UL33 gene; UL34 gene; UL35 gene; UL36 gene; UL37 gene; UL38 gene; UL39 gene; UL40 gene; UL41 gene; UL42 gene; UL43 gene; UL44 gene; UL45 gene; UL46 gene; UL47 gene; UL48 gene; UL49 gene; UL49A gene; UL50 gene; UL51 gene; UL52 gene; UL53 gene; UL54 gene; UL55 gene; UL56 gene; UL6 gene; UL7 gene; UL8 gene; UL9 gene; uracil-DNAP-glycosylase; US1 gene; US10 gene; US11 gene; US12 gene; US2 gene; US3 gene; US4 gene; US5 gene; US6 gene; US7 gene; US8 gene; US9 gene; US9 gene; US9 gene.

SOURCE human herpesvirus 2

ORGANISM Vitruvius dsDNA viruses, no RNA stage: Herpesviridae;
Alphanherpesvirinae; Simplexvirus.
1 (bases 1 to 154746)

REFERENCE McGeoch,D.J., Moss,H.W., McNab,D. and Frame,M.C.
AUTHORS DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome:
TITLE identification of the gene encoding glycoprotein G, and evolutionary comparisons

JOURNAL J. Gen. Virol. 68 (Pt 1), 19-38 (1987)

```

MEDLINE
REFERENCE
AUTHORS
TITLE
87111457
2 (bases 1 to 154746)
Everett,R.D. and Fenwick,M.L.
Comparative DNA sequence analysis of the host shutoff genes of
different strains of herpes simplex virus: type 2 strain Hg52
J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
90278430
3 (bases 1 to 154746)
McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.
Comparative sequence analysis of the long repeat regions and
adjointing parts of the long unique regions in the genomes of herpes
simplex viruses types 1 and 2
J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
92113549
4 (bases 1 to 154746)
Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J.
A novel herpes simplex virus gene (UL49A) encodes a putative
membrane protein with counterparts in other herpesviruses
J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
92356101
5 (bases 1 to 154746)
Dolan,A.
Direct Submission
Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church
Street, Glasgow, G11 5UR, UK
revised bu [6]
6 (bases 1 to 154746)
Dolan,A.
Direct Submission
Submitted (08-DEC-1999) A. Dolan, MRC Virology Unit, Church
Street, Glasgow, G11 5UR, UK
On Dec 19, 1999 this sequence version replaced gi:1869820.
Location/Qualifiers
1. 154746
/organism="human herpesvirus 2"
/strain="Hg52"
/db_xref="taxon:10310"
1. .254
/note="a sequence"
255
/note="start of TRL (terminal copy of Long Repeat region)"
440. .934
/feature="RL1"
/number=1
440. .1379
/feature="RL1"
join(440. .934,1089. .1379)
/feature="RL1"
/codon_start=1
/protein_id="CAB06759.1"
/db_xref="GI:1869821"
/db_xref="SWISS-PROT:P28283"
/translation="MSRRRGPRKRGPVRRRPGAPAVPRGAPVPRGALPPRDSOM
VPAIDSGTAVESAAPAAASLLRWLVPQADSDADADAGDDAEMANSPSEGGKAP
EAAHAAAPAAACPPPPPKERGPQPLPPLPHTLRITTEYLRLSLRRRPPSPAD
APGKQCFSPRVORHLVAVETARLARLGRSMRERDRPRFRRVAAAVIGPCLE
PEARARARARARHEDGCPAEEDEAAAAANGSSAAGPCRRAY"
446. .460
/rpt_family="7"
461. .547
/rpt_family="6"
943. .1070
/rpt_family="5"
1089. .1379
/feature="RL1"
/number=2
/product="neurovirulence factor"
1738. .1743
2303. .5368
/feature="RL2"
2303. .2377
/feature="RL2"

```

Q

[illegible]


```
JOURNAL      J. Gen. Virol. 68 (Pt 1), 19-38 (1987)
MEDLINE      87111457
REFERENCE    2 (bases 1 to 154746)
AUTHORS      Everett R.D. and Fenwick M.L.
TITLE        Comparative DNA sequence analysis of the host shutoff genes of
              different strains of herpes simplex virus: type 2 strain H652
JOURNAL      J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)
MEDLINE      90278430
REFERENCE    3 (bases 1 to 154746)
AUTHORS      McGeoch D.J., Cunningham C., McIntyre G. and Dolan A.
TITLE        Comparative sequence analysis of the long repeat regions and
              adjoining parts of the long unique regions in the genomes of herpes
              simplex viruses types 1 and 2
JOURNAL      J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)
MEDLINE      92113549
REFERENCE    4 (bases 1 to 154746)
AUTHORS      Barnett B.C., Dolan A., Telford E.A., Davison A.J. and McGeoch D.J.
TITLE        A novel herpes simplex virus gene (UL49A) encodes a putative
              membrane protein with counterparts in other herpesviruses
JOURNAL      J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)
MEDLINE      92356101
REFERENCE    5 (bases 1 to 154746)
AUTHORS      Dolan A.
TITLE        Direct Submission
JOURNAL      Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church
              Street, Glasgow, G11 5UR, UK
              revised by [6]
              6 (bases 1 to 154746)
REMARK       Dolan A.
AUTHORS      Direct Submission
JOURNAL      Submitted (08-DEC-1999) A. Dolan, MRC Virology Unit, Church
              Street, Glasgow, G11 5UR, UK
              On Dec 13, 1999 this sequence version replaced g1:1869820.
COMMENT      Location/Qualifiers
FEATURES     source
              1. 154746
              /organism="human herpesvirus 2"
              /strain="HG52"
              /db_xref="taxon:10310"
              1..254
              /note="a sequence"
              255
              /note="start of TRL (terminal copy of Long Repeat region)"
              440..934
              /gene="RL1"
              /number=1
              440..1379
              /gene="RL1"
              join(440..934,1089..1379)
              /gene="RL1"
              /codon_start=1
              /protein_id="CAB06759.1"
              /db_xref="GI:1869821"
              /db_xref="SWISS-PROT:P28283"
              /translation="MSRRRGPRRRGPRRRPGCAPVPRPGAPAPRRGALPTADSQM
              VPAVDSGTAVESAPASASLLRRLVPOADSDSDADYAGNDADAMANSPPSEGGKAP
              EAPAPAAACOPPPPRKRGPOPLRPHALRRTTETYLARLSLRRRRPPASPPD
              APKGVCFSPRVQYRHLVAMETARLARLRGSMARERDRDRFRFRVAAAEVIGPCLE
              PEAPARAKARAHEDGCPAEFEFAAAAGSSAAAGPGRRAV"
              446..460
              /rpt_family="7"
              461..547
              /rpt_family="6"
              943..1070
              /rpt_family="5"
              1089..1379
              /gene="RL1"
              /number=2
              /product="neurovirulence factor"
              1738..1743
              /protein_id="CAB06762.1"
              2303..5368
              /gene="RL2"
              2303..2377
              /translation="MFSASVTPPOPLGLSDATNPPLPTSVPLDMAAFRAFLIDDARW

CDS          join(2303..2377,2785..3462,3644..5368)
              /gene="RL2"
              /number=1
              /protein_id="CAB06760.1"
              /db_xref="GI:1869822"
              /db_xref="SWISS-PROT:P28284"
              /translation="MEPRGTSRADGPRPPRRPGRPGTNPAPAHAMGNDMOWLAS
              SDSEETEVGTSDDLDHRDTSRAGSTDTMFAGLMDATPPAPPAEKGSPFPAD
              AGSCGGGPVGEAEAEAGGGDVCACVTDIAPLRQSPCLHPFQIPMKWVPLR
              NTPDLCTPVALVIVAGTSFSTIPYVNDPRTVREAEAVRAGVADITWGNPRT
              APKSLSGHTVVALSPTPPWPGTDDDDLDADYVPPAPRRARRGGAGATRG
              SOPAATPPAPPGAPRSSSGGAPLRAGVGSGGAPVAVVPRVAVSLPPAAGGRRO
              ARVGEDEMAABERTPPAROPRAOEPPIYISGPPSPRRAPGPGTSSSSAOY
              SSGPGGGLPQSSGRRAPRAAIVPRSPRAAIVSASADAGAPAPVVDH
              RAPRSRMTQAOQDITQASLIRAGATDARGSGGAEAGPGVPGTNPAPPAEGLA
              AAPRRKRSGDSGPAASSASAAAPRSPPLAPGVGAKRAAPRAPDSDGDRHGFL
              APASAGAAPSPASPSOAAVAASSSASSSASSSASSSASSSASSSASSSA
              SSSAGGAGGSVASAGGERRTSLGPRAAPRGPKCAKRTTHAEGGPRGRDAP
              GTRIRLPACVSSVALAPYVNTYTGDCIPYIDMETHGATVLYVDQGNADLIR
              AAPAMSRRTLPEHARNCVPPDIPTPPASEMNSIMTIVGNLFDQGLVGLDHP
              GLRSRHPWSREOGAPAPADAPAGHE"
              2785..3462
              /gene="RL2"
              /number=2
              3468..3607
              /rpt_family="4"
              3644..5368
              /gene="RL2"
              /number=3
              /function="modulator of cell state and gene expression"
              4655..4771
              /rpt_family="3"
              /complement(4811)
              /note="proposed LAT splice acceptor site"
              5618..5623
              /complement(7025)
              /note="proposed LAT splice donor site (5' end of stable
              LAT species)"
              /complement(7732)
              /note="primary transcription initiation site for latency
              associated transcripts (LAT's)"
              9047..9193
              /rpt_family="2"
              9276..9297
              /rpt_family="1"
              9298
              /note="start of UL (Long Unique region)"
              9427..10101
              /gene="UL1"
              9427..10101
              /gene="UL1"
              /codon_start=1
              /product="viroion glycoprotein L"
              /protein_id="CAB06761.1"
              /db_xref="GI:1869823"
              /db_xref="SWISS-PROT:P28278"
              /translation="MGFVCLFGVYMGAMGAMGSGQATEYVLSYIAKEVGDILRVPC
              MRPADDSVSWRYEAPSVIDYARIDGILRYHCGGLDFLDRAQRAQVYNPLFAAG
              FLBDLSHVSVPADTOETTRALYKEIRDLGSRKQVSHAPVACGVNDVYSRTRC
              VGRDLRPANTTSTWEPVSSDDEASQSKPLATOPPVLAISNAPRRRVSPTGRRRH
              TLRARN"
              10211..10978
              /gene="UL2"
              10211..10978
              /gene="UL2"
              /codon_start=1
              /product="uracil-DNA glycosylase"
              /protein_id="CAB06762.1"
              /db_xref="GI:1869824"
              /db_xref="SWISS-PROT:P28275"
              /translation="MFSASVTPPOPLGLSDATNPPLPTSVPLDMAAFRAFLIDDARW
```


Mon Oct 16 10:08:43 2000

us-09-431-843b-13.rge

Page 18

[illegible]

Search completed: October 12, 2000, 19:05:23
Job time: 38899 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 17:56:51 ; Search time 17897.3 Seconds
(without alignments)
390.701 Million cell updates/sec

Title: US-09-431-843b-7
Perfect score: 1601
Sequence: 1 tagaatcagcgccgcctga.....ccctctcgagcgtggtggcg 1601

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenDbml: *
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pl4:*
11: gb_pl5:*
12: gb_pl6:*
13: gb_pl7:*
14: gb_pl8:*
15: gb_pl9:*
16: gb_pl10:*
17: gb_pl11:*
18: gb_pl12:*
19: gb_pl13:*
20: gb_pl14:*
21: gb_pl15:*
22: gb_pl16:*
23: gb_pl17:*
24: gb_pl18:*
25: gb_pl19:*
26: gb_pl20:*
27: gb_pl21:*
28: gb_pl22:*
29: gb_pl23:*
30: gb_pl24:*
31: gb_pl25:*
32: gb_pl26:*
33: gb_pl27:*
34: gb_pl28:*
35: gb_pl29:*
36: gb_pl30:*
37: gb_pl31:*
38: gb_pl32:*
39: gb_pl33:*
40: gb_pl34:*
41: gb_pl35:*
42: gb_pl36:*
43: gb_pl37:*

Result				SUMMARIES			
No.	Score	Query Match	Length DB	ID	Description		
1	1601	100.0	1676	11	AF172450	AF172450 Homo sapi	
2	1352.6	84.5	2423	11	AF172451	AF172451 Homo sapi	
3	1352.2	84.5	2363	11	AF172452	AF172452 Homo sapi	
4	1349.4	84.3	2483	11	AF172453	AF172453 Homo sapi	
5	1269.2	79.3	2423	11	AF109134	AF109134 Homo sapi	
6	979	61.1	1232	11	AF172449	AF172449 Homo sapi	
7	699.8	43.7	160241	35	HS88517	AF103565 Human DNA	
8	686.6	43.5	2208	12	AF156878	AF156878 Rattus no	
9	693.4	43.3	6670	11	AF112980	AF112980 Homo sapi	
10	72.8	4.5	7218	5	I66494	I66494 Sequence 14	
11	66.4	4.1	236	91	HSRMC2C3	X76589 H.sapiens D	
12	65.4	4.1	214721	46	AC024957	AC024957 Mus muscu	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Db 481 AGCOTCAGGAGGTCGAGGTGTTAAAGCTCCACAGAGATCCAGAGCGGCTTGCCG 540
Qy 541 gcttaagagctcatgtctgggtcttctacggatccggtgtaggaacgagacgagcgag 600
Db 541 GCTTACAGAGTCTATGCTGGGCTTCTACGGGATCCGGCTGAGAGACGAGGACGCGCAG 600
Qy 601 gtgggcccagacagaaactaccagaagcgcttccagaaacttgagctggcgagccagaac 660
Db 601 GTGGGCGGAGACAGAACTACAGAAAGCGCTTCCAGAACTGTAAGTGGCCAGCCACAC 660
Qy 661 aacctccgacacacagcatctcctcaagtcgctgtgtagctgagcctcgacatccag 720
Db 661 AACCTCCGATCAACGACATCTCAAGTCCGCTGTAGTGAAGCTCGAGACTTCCAG 720
Qy 721 ggcgcactgtctcgtcttctcctggagagacgtgtgtggcgagagctgcccgggggtg 780
Db 721 GCCCACTGCTGCTCTTCTCTGAGAGACGCTGTGGGGAGCTGCGGGGGGTG 780
Qy 781 cggcagagtgcccttggaactatctatctgctgctgctgctgctgctgctgctgctg 840
Db 781 CGGACAGTCCCTGAGACTTCTCATGTTGCGCGTGCCTGCGGACACAGCCGCCGAC 840
Qy 841 ctgtgtcaactgctgctgtaggaactccggtccgctgcaagtctgtgtggggccca 900
Db 841 CTGTGTCACTTCCCTGAGGACACTTCCGGCCCGCTGCAAGTCTGTGGGGCCCAA 900
Qy 901 gacaactgagaggttcaagacccactctctgctgctgctgctgctgctgctgctgctg 960
Db 901 GACAACTGCGAGGTTCAAGCCCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Qy 961 gtggagagagaaagaccccgaggaccccgagacacagagcgagaccccgaggtctgagc 1020
Db 961 GTGAGAGAGAAAGACCCCGGAGACCCGAGACACAGAGCCAGCCAGGGTGGAGAC 1020
Qy 1021 tctgggcccagagatgtagaggtgtgggtgtagaggtgtagaggtgtagaggtgtag 1080
Db 1021 TGTGGGCCAGAGCATAGCAAGGTTGGGGTGGAGGCTGAGAGGGGCCACAGCCAGC 1080
Qy 1081 gtggaaccccaagatgcgaggaacccctgtgagagagcagagggatgtaggtagggggc 1140
Db 1081 GTGGAACCCCAAGATGCGGAGACCCCTGAGAGAGAGCCAGGGGATAGGCAAGGGGCC 1140
Qy 1141 ggggaagatagcccgagaccccttaagcccccagaagagagagagagagagagagctg 1200
Db 1141 GGGGAAGATAGGCCGAGCCCTTAAGCCCAAGAGAGCAAGAGAGAGAGAGAGAGAG 1200
Qy 1201 agccggcgagagcagcgccgcccacagagcagagccctcagaagtgtcagaagttag 1260
Db 1201 AGCCGGCGGAGAGCGCCGCCACAGAGCCAGGCTTCAAGAGTGCCTCAAGAGGTGAG 1260
Qy 1261 atcgctctgaatttgagaggtgtgtgcccacagcagagcagcagcagcagcagcagc 1320
Db 1261 ATCGCTCTGAATTTGAGAGGGGTGTGCCCTCAGCCAGGAGAGCTCAGAGCGGAGC 1320
Qy 1321 gaagtgtgagtcagagacccctggagagcctctcctgtcagaggggtgtggggctc 1380
Db 1321 GAAGTGTGAGTCAGAGACCCCTGGGAGGAGTGAAGCCCTGCCCAACCCCTGGAGGC 1380
Qy 1381 cgg 1383
Db 1381 AGG 1383

RESULT 3
AF172452
LOCUS AF172452 2363 bp mRNA PRI 19-Apr-2000
DEFINITION Homo sapiens oploid growth factor receptor mRNA, complete cds.
ACCESSION AF172452
VERSION AF172452.1 GI:7595304
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
1. 2363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue="placenta"
34. 2007
/note="[Met-5]-enkephalin binding protein; OGR;
alternatively spliced"
/codon_start=1
/product="oploid growth factor receptor"
/protein_id="AA64405.1"
/db_xref="GI:7595305"
/translation="MDPPDCDSWEDEDEDAEDAEDEDECEDEGAEAGARADADEDEE
SEEPRAAPSSFQSRMTRATRMCMYRNYRNPDLVREDNGTPTNLSFYRNEIR
FLPNCFIEDILNMTDNYDLDDNHSYIOMLPFRPGVNMWAKPLTLEUEVFRS
OEIERLVRAVELMGEYIGTILRDGRTGAGRAONOKRONLNMRSNHNLRTRTLK
SPECILSHEHOAPLVREPLEETLVRETPVRSALDIPFAYRCHOROLYHME
HPRRCFTWGPDKLRKRPSSLPPLBESKRVVEESGPDHASTOGRTGPEH
SKGGRVDESPQSVEPDAGPLENSQGEAGHGHDREPSPKSKRLULSR
KQEPTEGPGSAEVERIALNLEGCALSQSLRTGTQEVGGDPGSAVPCROPLSR
VADKVRKRRVDEGASAAVAGAGATLALASPAKSPKAGHSENVEDTEGRT
PKPEGTGSPSETPGSPAGPADDEPAESSETPGSPAGPADDEPAESSETPGSP
AGPRDEPAESSETPGSPAGPADDEPAESSETPGSPAGPADDEPAESSETPGSP
SPAGPRDEPAKGEAELODAVBSASGKR"
BASE COUNT 488 a 729 c 825 g 320 t 1 others
ORIGIN
Query Match 84.5%; Score 1352.2; DB 11; Length 2363;
Best Local Similarity 98.6%; Pred. No. 1.1e-191;
Matches 1363; Conservative 1; Mismatches 19; Indels 0; Gaps 0;
Qy 1 tagaattcagcgccgctgaattctagccgagcatgtagcagaccgagctcgactccacc 60
Db 1 TAGAATTCAGCGCGCGCTGAATTTAGCCGAGCATGAGACGCCGACTCGACTCCACC 60
Qy 61 tggagagagagcgagagagatgtagagagcagagagacgagacgtagcagagcgag 120
Db 61 TGGAGAGAGAGCGAGAGAGATGCGAGAGACGCGAGACGAGACTCGAGAGCGGAG 120
Qy 121 gccgcgcgcgagagagcagcagcagagagagagagagagagagagagagagagagag 180
Db 121 GCCCGCGCGGAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy 181 gcggcgagcccaagctcgttcacagtcagaatgacagaggtccagaaactgagcagcag 240
Db 181 GCGGCGGCGGAG 240
Qy 241 agggacatgtgtagatctggcacacatccggatctgtgtgagagagagacgtagatgg 300
Db 241 AGGACATGTGTAGATTCGGCACACACTATCCGATCTGTGTGAGAGAGAGACTCAATGGG 300
Qy 301 gacacgccaacactgagttctacagaaatgagatccgctctgtcccaacgagctgttc 360
Db 301 GACACGCCAAACCTGAGTTTCTACAGAAATGAGATCCGCTTCTGCCCAACGCGCTGTTT 360
Qy 361 attgaagacattcttcgaactgagcagcagacatgacctcttgaagacaactacc 420

D	361	ATTGAGGACATTCTTCAGAACTGACGACAACTATGACCTCTTGAGGACATCACTTCC	420
O	421	tacatccagatggtctgttctctctgagagacagagatgaactgcatgccaagccctc	480
D	421	TACATCCAGATGAGTGTCTTCTCTGCGAGAACGAGGATCACTGGATGCCAAGCCCTC	480
O	481	acgcctcaagaggttcgaagtgcttaaaagctccacaggaatcacagagcgctgtccg	540
D	481	ACGCCTCAGGAGTGTGAGGTGTTTAAAACTCCAGGATATCAGAGGCGCTTGCCGG	540
O	541	gctcagcagatcgtctggtgtcttctacagagatcgctgtgagagaccagagcgagc	600
D	541	GCTCAGAGCTCATGCTGGGCTTCTTACGGGATCCGGCTGAGGAGCCGAGGCGGCGAG	600
O	601	gtggcgagcagcacaagaactcagagcgcttcacgaactgaactggcgagccacaac	660
D	601	GTGGCGCAGCAGACAGAACTACCAAGAGCCGCTTCCAGAACTTAACCTGCGGCGCAAC	660
O	661	aacctccgcatcacacgcactcctcaagtgcggtgtgaactgagcctcagacattccag	720
D	661	AACCTCCGATACACAGCATCTCTCAAGTGGCGGTGTGAGCTAGACCTCAGCATTTCCAG	720
O	721	gcgcacactggtcgtcttctctctgagagagacgctgtgtgagcgagagctgcggggtg	780
D	721	GCGCAGCTGTGCTCGGCTTCTCTGAGAGAGAGCTGTGTGCGGCGGAGCTGCCGGGGGTG	780
O	781	cggcagagatgacctgactactcatgttctgcggtgtgctgtgcagacagcgccgacg	840
D	781	CGGCAGAGTGGCTTCTGAGACTTCAATGTGCGCGGCTGCCGACACAGCGCCGCCAG	840
O	841	ctgggcaactgcctgtgagagacattccgcccgcctgtgaagtctgtctgggggcccac	900
D	841	CTGGTGCACTTGCTGCGGAGACATTCGCGCCCTCCGCTGCAATTTGCTGTGGGGGCCCA	900
O	901	gacaaagtgcgagaggttcaaacccagctctctgcgcgacatcgctcgaggggtctcagaa	960
D	901	GACAAAGTGCAGGAGTTCAAGCCAGCTCTGTGCCSANTCCGCTCGAGGGGTCTCAGGAAG	960
O	961	gtggagggag	1020
D	961	GTTGGAGGAGGAGGAG	1020
O	1021	tggtggcagagcagcag	1080
D	1021	TGTGGGCCAGAGACATAGCAAGAGGTGGGGGCAAGGTGAGAGAGAGAGAGAGAGAGAG	1080
O	1081	gtggagccccagagatgcgagagagagagagagagagagagagagagagagagagag	1140
D	1081	GTTGAGCCCCAGAGATGCGGAGCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1140
O	1141	ggggagagatagcgag	1200
D	1141	GGGGAAGATAGAGCCGAGAGCCCTTAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG	1200
O	1201	agccgagcag	1260
D	1201	AGCCGCGGAGAGACAGCCGCCACAGAGCCAGCCCTCAAGAGTGTGAGAGAGAGAGAG	1260
O	1261	atcgctctgaatttggag	1320
D	1261	ATCGCTCTGAATTTGAGAGGGGTGTGACCTTCAAGCCAGGAGAGCTCAGAGAGAGAGAG	1320
O	1321	gaagtgagcgagcag	1380
D	1321	GAAAGTGGGGGTGAG	1380
O	1381	cag 1383	
D	1381	AGG 1383	

RESULT 4 AF172453 .

LOCUS	AF172453	2483 bp	mRNA	PRI	19-Apr-2000
DEFINITION	Homo sapiens	opioid growth factor receptor mRNA, complete cds.			
ACCESSION	AF172453				
VERSION	AF172453.1	GI:7595306			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.				
TITLE	1 (bases 1 to 2483)				
JOURNAL	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.				
MEDLINE	Cloning, sequencing, chromosomal location, and function of cDNAs encoding an opioid growth factor receptor (OGFR) in humans				
REFERENCE	Brain Res. 856 (1-2), 75-83 (2000)				
AUTHORS	20143540				
JOURNAL	2 (bases 1 to 2483)				
REFERENCE	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.				
AUTHORS	Direct Submission				
JOURNAL	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA				
FEATURES	Location/Qualifiers				
SOURCE	1. 2483				
CDS	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="20"				
	/map="20q13.3"				
	/tissue_type="placenta"				
	34. .2127				
	/note="(Met-5)-enkephalin binding protein; OGFR; alternatively spliced"				
	/codon_start=1				
	/product="Opioid growth factor receptor"				
	/protein_id="AA64406.1"				
	/db_xref="GI:7595307"				
	/translation="MDPDCSTWEDEEDAEADDEDECEDEGAAGADADGDEDE SEEPRAAPSPFOSRMTRMCRATRMCRHYPPLVDEDCDGTDPNLSFVNEIR FLVNGCFIEDILQWNTQNYDLEENHAYIQLFLRPGVNRHAKPLTLEVEYFKS OEIORDERVATLEMLGFTGLERDGTGIVGRAQNYKRONLWNRSHNDRTILK SPCELSLEHROAPLVRELETLERGLPVROSLDYFMFAVCROROLVHFAWE HFRRCRFEVWGPQDKLRFKPSLPHLEGRKVEEGSGPDPHEASTGRICEPH SKGGRVDEGPORSVEPDOPGLERSQDEAGHGDEPREPLSPKSKRKLISR EOPPTGGPOSASEVERIALNLEGCALSTSLRTGQEVGODPEAVOCROPLGAR VADKVRKRRKTDSTGSAVYASGATLTLAGSPARSHPKASHNSNGVEDPGRTP GKKEGTGSPSEITGPSPAGPAGDEPKTSETGSPAGTROPESSESTGPAP AGPAGDEPAESPESTGPSPAGPAGDEPAIPSETGPSAPGTRDPASPSTGP RPAPADDEPAESPESTGPSPAGPAGDEPAESPESTGPSAPGTRDPAPKGEAAE LQDAEVSSAKSKP"				
BASE COUNT	518 a 778 c 861 g 326 t				
ORIGIN					
Query Match	84.3%; Score 1349.4; DB 11; Length 2483;				
Best Local Similarity	98.5%; Pred. No. 2.8e-191;				
Matches 1362; Conservative	0; Mismatches 21; Indels 0; Gaps 0;				
O	1 tagaattcagcgccgctgaattctagccgagcatgtagcagaccgagctgcagctcac	60			
D	1 TAGAATTACGCGCGCGCTGAATTTCAAGCCGAGCATGAGACAGCCCGACTGCGACTCCAC	60			
O	1 tgggag	120			
D	1 TGGGAG	120			
O	1 ggc	180			
D	1 GCCCGCGCGCGAGGAGAGCGGAGCGAGGAGCGAGAGAGAGAGAGAGAGAGAGAG	180			
O	1 gc	240			
D	1 GCG	240			
O	241 agggacatgtatagtcacacatactccggtctgtgtgagacagcagctgcagtg	300			

Db 241 AGGACATGCTAGTATCGGACAACTATCCGATCTGTGAGACAGACTGCAATGGG 300
QY 301 gacacgcaaaccttgatcttctacagaatgagatccgctctctgcccacggtcttcc 360
Db 301 GACACGGCAAACTGATTTCTACAGAAATGATCCGCTTCCGACCAAGGCTGTTTC 360
QY 361 attgagacacttctcgaacttgacgacacatactactcctcttgagacaactacc 420
Db 361 ATTGAGGACATTTCTTCAAACTGGACGGAACAACTATGACTCCTTGAGCAATCACTCC 420
QY 421 tacatccagtggtcttctctctcgcgagacagagtgaaacttgacatgcaatgcccctc 480
Db 421 TACATCCAGTGGCTTCTCTCTGCGAGAACAGAGATGAACTGGCATGGCAAGCCCTTC 480
QY 481 acgctcagggaggtcgaggtgtttaaaagctcccaagagatccagagagcggtctgtccg 540
Db 481 ACCGTCAGGAGGTCGAGGTGTTTAAAGCTCCAGAGATCCAGAGAGCGGCTTGTCGG 540
QY 541 gacctacagatcgtctggtctcttaagagatccggtctgagacagagcgagcgag 600
Db 541 GCGTACAGCTCATGCTGGGCTTCTACGGGATCCGCTGAGAGACGAGGACGAGGCGACG 600
QY 601 gtggccgagcagacagactaccagaaagcgcttcacagaaacttgagcgagcgacacac 660
Db 601 GTGGCGGAGACAGAACTACAGAAAGCGCTTCAGAACTGAACTGGCGACCAAC 660
QY 661 aacctcgcacacacagcatctcctcaagtcgcgtgtgagctgagcctcgacatccag 720
Db 661 AACCTCGGCTATCACAGCACTCAAGTCCGCTGTGAGTGAAGCTCGAGCTTCCAG 720
QY 721 ggcacactggtcgtctctctctctgagagagcgtgtgagggagagctgagcgaggtg 780
Db 721 GCCCACTGCTCGCTTCTCTCTGAGAGACCTGCTGCGGAGGCTGCGGCGGAGG 780
QY 781 cggcagagtgcccttgactactcaltctgctgctgctgctgctgctgctgctgctgctg 840
Db 781 CGGCAAGTGCCTCGGACTACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
QY 841 ctgtgtcacttgcgtcgtgagacactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 841 CTGTGCTACTTGCCTGAGGACCTTCCGCGCCGCTGGAAGTCTCTGAGGCGCCCA 900
QY 901 gacaaagctcggaggtcacaagcccaactctctgctgctgctgctgctgctgctgctg 960
Db 901 GACAAAGCTCGGAGGTTCAAGCCCACTCTCTGCGCATCCGCTGAGGCGCTCCAGAGAG 960
QY 961 gtggag 1020
Db 961 GTGAG 1020
QY 1021 tctgag 1080
Db 1021 TCTGAG 1080
QY 1081 gtgag 1140
Db 1081 GTGAG 1140
QY 1141 gggag 1200
Db 1141 GGGAG 1200
QY 1201 agccgag 1260
Db 1201 ACCCGGAG 1260
QY 1261 atcgcctcgaattctgag 1320
Db 1261 ATCGCTCGAATTTGAGAGGAGGTGCTCCCTACAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 gaagtgag 1380
Db 1321 GAAAGTGGCGGTCAGAGACCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

Db 1321 GAAAGTGGCGGTCAGAGACCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 cgg 1383
Db 1381 AGG 1383
RESULT 5
AF109134 2423 bp mRNA PRI 10-JAN-1999
LOCUS Homo sapiens 7-60 mRNA, complete cds.
DEFINITION AF109134
ACCESSION AF109134
VERSION AF109134.1 GI:4139227
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2423)
AUTHORS Liu, J., Brewton, R.G., Takanosu, M., Wood, B.M. and Mayne, R.
TITLE Cloning of 7-60: A human gene from chromosome 20q13.3 which encodes
a novel intracellular protein motif repeated seven times
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2423)
AUTHORS Liu, J., Brewton, R.G., Takanosu, M., Wood, B.M. and Mayne, R.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1998) Cell Biology, University of Alabama at
Birmingham, 1670 University Blvd. VHS02, Room 605, Birmingham, AL
35294, USA
FEATURES
source location/Qualifiers
1..2423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/note="approximately 3kb upstream of COL9A3"
206..2062
/codon_start=1
/product="7-60"
/protein_id="A003737.1"
/db_xref="GI:4139228"
/translation="MTGSRMWRATRMCRVRYHNPDLVERDCNGDPTNLSFYRNEIR
LPNGCFTEILQWNTDNYLDLNDNSHYIOMFLPRPGVNMHAKPLTILEVEYFKSQQ
EIQPRVYRAAHAGLIRDPAGSPRHHGSPSTLEALPELPRSHNNLRTTRILKS
LGELGHEHPALVREFLESILVRRLPGVPSALDYFMFVAGCRHRORLVFAWES
FRPRCKFWGPDOKLRFKPSLSLPHLEGRKVEEGSGDPDHEASTGRTGCPHS
KGGRVDEGPOPRSPVODAGPLERSQGEAGGHEGDRPREPLSPKSKRXLELSRE
OPTREGPQSASEVERIALNLCECALSGSLIRGTGQEVGGDGPGEAVOPCRPIARV
ADKVRKRRKYDEGASAVASGAGTALAGSPAPAGDEPRAPAGDEPRAPAGDEPRAPAG
PKRGTSGPSSETGSPAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAG
GPADEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAG
PAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAGDEPRAPAG
ORIGIN
BASE COUNT 501 a 759 c 846 g 317 t
Query Match 79.3%; Score 1269.2; DB 11; Length 2423;
Best Local Similarity 97.3%; Pred. No. 2; e-179;
Matches 1322; Conservative 0; Mismatches 33; Indels 3; Gaps 3;
QY 27 gccagcatgacgagcagcccgactcgactcactcgtagagagagagagagagagagagagag 86
Db 23 GCCAGCATGAGACGACCCCACTGCGACTCCACTTGAGAGAGAGAGAGAGAGAGAGAGAG 82
QY 87 ggaagcggag 146
Db 83 GGAAGCGGAG 142
QY 147 agggag 206
Db 143 AGGGAG 201
QY 207 cagagatgacaggttccagaaactgagcagagcagagagacatgtatgagtagtcagacaa 266

|||||
Db 202 CAGATGACAGGCTCCAGAAACTGGCGAGCCAGGAGACATGTGTAGTATCGGACAA 261
QY 267 ctatccggatctgtgtgaaagagatgcaatggtggagacccaaactgagttctacag 326
Db 262 CTATCCGGATGTGGTGGACAGACTGTCAATGGGGACCGCAAACTGAACTTTCTACAG 321
QY 327 aaatgagatccgtcttcctgcgcgaacgctgtttcatatgaaagacattcttcaagatgac 386
Db 322 AAATGAGATCCGCTTCCCTGCCCAAGCGCTGTTCATTAGAGCATTTCTTCAGAACTGGAC 381
QY 387 ggaacaataatgacctccttgaagaaataactctacataccagtggtcttccctgcy 446
Db 382 AGACAATATGACCTCCTTGAGGACAAATCACTCCTACATCCAGTGGCTGTTCCCTGCG 441
QY 447 agaaacaggagatgaaatgcatgccaagccctcaacgctcaaggagatgcgagtggttaa 506
Db 442 AGAACCAGAGTGAACCTGGCATGCCAACCCCTCACGCTCAGGGAGGTGAGGTTTAA 501
QY 507 aagctcccaagagatccagagagcgctgtctcgagccctacg-aggctatgcttgagcttct 565
Db 502 AAGCTCCAGAGATCCAGAGCGGCTGTCCGGCTTACCGACTCATGCTGGGCTTCT 561
QY 566 accggaatccgctggaagagaccgagccaggtgagcaggtgagcagacagactacaga 625
Db 562 ACGGATCCGGCTGGAGAGACGAGCAGGGGACGGTGGCGGACAGACAGAACTACAGA 621
QY 626 agcgcttccagaactgaaactgagcgagccagccaaacaaactccgcatcaagcatctca 685
Db 622 AGCGCTTCCAAACCTGAACT-GGCGACCCCAACACCTCCGATCACAGCATCTCCCA 680
QY 686 agtcgcgtgtgagctgagctgagcacttccagcgccactgctcgtcttctctg 745
Db 681 AGTGGCTGGGTGACTGGGCTTCCAGCATCTCCAGCGCCCTGGTCCGCTTCTCTGG 740
QY 746 aggaagacgtgtgtgagcgagagctgcccgggtgtgagcagagtgcccttgactactca 805
Db 741 AGGAGAGCGCTGTGCGGGGAGCTGCGGGGTGCGGAGAGTGGCCCTGGACTACTTCA 800
QY 806 tttctgcgcgtgctgtcgcgaacacagcgccgacgctgtgtcaattcgcttggagcaact 865
Db 801 TGTTCGCGGTGGGTGCGGACACGCGCCCGACGTGTGTACATCTCCCTGGGAGCACT 860
QY 866 tcgcgccccgctgcaagtgtctgtgggggccccaaagaaactgcgagaggttcaagccca 925
Db 861 TCCGGCCCCGCTGCAAGTCTGTGGGGGCCCAAGACAACTGGGGGTTCAAGCCCA 920
QY 926 gctctctgcgcacatccgctcagaggtctcagaaagtgagaaagaaagaaagcccgagg 985
Db 921 GCTCTGTGCGCATCCGCTCAGAGGCTCAGAAAGGTGAGAGAGAGAAAGAACCCCGGG 980
QY 986 accccgacacagagccagaccccaagggtcgagacctgtggccagagcatatgcaagggtg 1045
Db 981 ACCCGGACCAAGAGGCGCCAGCACCGAGGCTCGACCTGTGGCCACAGCATAGCAAGGCTG 1040
QY 1046 ggggcaaggtgtaagagggggcccccaagccagagcggtgaaagcccccaagatgcggagccc 1105
Db 1041 GGGGCAAGGTTGAGAGGGGGGCCCAAGCCAGAGCGAGGTGAGACCCCAAGATCGGGACCC 1100
QY 1106 tggagagagagcaggggagatagagcagggggccagcggaagatagagccggagccctaa 1165
Db 1101 TGGAGAGAGAGCCAGGGGATAGAGCAGGGGCCAAGGGGAAGATAGCGCGGAGCCCTTAA 1160
QY 1166 gccccaagagagcaagaagaagaagtgtgagctgtgagccgagcggaagcagccagccag 1225
Db 1161 GCCCCAAGAGAGCAAGAAAGAGAACTGTGAGCGCGGAGAGCAGCCGCCCAAG 1220
QY 1226 agcagagccctcagagtgccctcagaggtgagagagatgcctgaatttgagaggggtgtg 1285
Db 1221 AGCGAGGCGCTCAAGTGTCTCAGAGGTGTGAGAAAGATGCTCTGAAATTTGAAGGGGTGTG 1280
QY 1286 cctcagcagagagcctcagagacgggagccagaaagtgtggcgctcagagccctgggg 1345
|||||

Db 1281 CCTCAGCAGAGGCGAGCCTCAGACGGGAGACCAGAAAGTGGGGGTGAGAGACCTGGGG 1340
QY 1346 aggcctcctgtccctgtcgcagggaggtggtggtcccg 1383
Db 1341 AGCGAGTGACACCTGCGCCACACCTCGGAGGCCAGG 1378

RESULT 6
AF172449
LOCUS AF172449 1332 bp mRNA PRI 19-APR-2000
DEFINITION Homo sapiens clone 127 oploid growth factor receptor mRNA, complete cds.
ACCESSION AF172449
VERSION AF172449.1 GI:7595298
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1232)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J. Cloning, sequencing, chromosomal location, and function of cDNAs encoding an oploid growth factor receptor (OGFr) in humans Brain Res. 856 (1-2), 75-83 (2000)
2 (bases 1 to 1232)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J. Direct Submission
Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA
LOCATION/Qualifiers
FEATURES
source
1..1232
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/clone="127"
/tissue-type="placenta"
34..1212
/note="[Met-5]-enkephalin binding protein; OGFr; alternatively spliced"
/codon_start=1
/product="oploid growth factor receptor"
/protein_id="AA64402.1"
/db_xref="GI:7595298"
/translation="MDDPDCDSITWEDEDAEDDEDEDEGEAAGADADAGDEDESEPPRARPSFSQRTGSRNMRATRDRCRYRNYPDLYERDNGDTPNLSFYRNRIRFLNPGFIEDIIONMTDNDYDLIEDNSYIOWLEPLREPVNNAKRLTIREVYFSSQETQDELYRAVEMLGFTGIRLEDRTGYVRAQNYQKRFQNLNMSHNRLRTTRILKSPCLSLHFOAPLVRFLEETLVRELEGEVROSALDYFMFAVCRHQRLVHFAWEHFRPRKFTWGPQDKLRKPSLPLRLGSSRVEEGAPGDEPASPSPAGPTRDEPAKAGEAEACCLAVSSHPLPCAPVFVPRPWSGRRAGLAFLSLMPSKAFSE"
BASE COUNT 256 a 369 c 403 g 204 t
ORIGIN
Query Match 61.1%; Score 979; DB 11; Length 1232;
Best Local Similarity 99.5%; Pred. No. 3e-136;
Matches 982; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tagaatcagcggcgcgtgaattcttagcagcatatgacgaaccccgactcgactccacc 60
Db 1 TAGAATTCAGGCGCGCTGAAATTCAGCATTCAGCAGACCCCGACTCGACTCCACC 60
QY 61 tggagagagagcagagagaggttgcgagagcgcgagagagcagagactgcgagagcgag 120
Db 61 TGGAGAGAGAGCGAGAGAGATGCGAGAGACGGAGGACGAGCATCTCGAGAGCGCGAG 120
QY 121 gccgcgcgcgcgaagagacgcagcagcgagggagacagagagatctcgagagagcgcg 180
Db 121 GCCCGCGCGGAGAGGAGACCGGAGACGAGGGAGCAGAGACGAGAGTCTGAGAGACCGCGG 180
|||||

```

QY 181 gggggcgccgcagctcgttccagtcgaatgacaggtgccagaatggcgagccag 240
    |||
Db 181 GGGGGCGCCCGCAGCTCGTCTCCAGTCCAGATACAGAGGGGCCAGAACTGGCCAGCCAG 240
QY 241 agggacatgttagatcagccacaataccgagctgtgtgaacagactcgaatgg 300
    |||
Db 241 AGGACATGTGTAGTATGGGACAACTATCCGATCTGGTGAACGACACTGCATGGG 300
QY 301 gaacagccaaactgtgttctacagaatgataccgtctcctcgaagcgtgttcc 360
    |||
Db 301 GACACCCCAAACTGAGTGTCTACAGAAATGAGATCCGCTTCTCCGCCAACGGCTGTTC 360
QY 361 atgaggaactcttcagacactgaagcgaactatgactccttcgaagaatcctcc 420
    |||
Db 361 ATGAGGAACTTCTTACAGACTGACGACAACTATGACCTCTTGAAGCAATCACTCC 420
QY 421 tacatccagtgctgttctcctctgcagagacagagatgaactgtcatgccagccctc 480
    |||
Db 421 TACATCCAGTGGCTGTCTCTCTGCGAGAACGAGTGAAGTGCATGCCAAGCCCTTC 480
QY 481 agcctaaagaggtcgaagtggttaaaagctcccaagagatcccaagcgcgttgcgg 540
    |||
Db 481 AGCCTAAGGAGGTGAGGTGTTTAAAGCTCCACAGAGATCCAGAGAGCGGCTTGTCCGG 540
QY 541 gctacagagctcatgtgtggtctctacagagatccggtcagagacagcagcagcag 600
    |||
Db 541 GCTACAGACTCATGCTGGGCTTCTTACGGGATCCGGCTGAGAGACCGAGCCAGGGCAG 600
QY 601 gtggcgacagcaagaactaccagaagcgtctccagaacctgaactgtgcgcacacac 660
    |||
Db 601 GTGGCGCGGACAGACAGAACTACCAAGAGCGCTTCCAGAACTGAGCGGCGCACAAAC 660
QY 661 aacctccgatacagcagctcctcaagtgcgcgtgtgagctgaagctcgagacttccag 720
    |||
Db 661 AACCTCCGATACACAGCACTCTCAAGTGCCTGTGACCTAGCCTCGACACTTCCAG 720
QY 721 ggcgcactgtgcgtctctctctcgtgagagacgtgtgcgcgcgcgcgcgcgcgcgc 780
    |||
Db 721 GCGCCACTGTGCTCGGCTTCTCTGAGAGAGAGCGTGTGCGCGGAGAGTGCCTGGGGG 780
QY 781 cggcagagtgcccttggaactacttcatgttcgcgtgtgcgtgcgcagacacgcgcgcag 840
    |||
Db 781 CCGCAGAGTGCCTTGGAATCTTCAATGTTCGCGCTGCCGACACACGCGCGCCAG 840
QY 841 ctggtgcaactgcgtgtgagacacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
    |||
Db 841 CTGGTGCACTTGCCTGGAGACACTTCCGCCCGCTGCAAGTGTCTGGGGGCCCCAA 900
QY 901 gacaaagtgcgaggtlcaagcccaagctctgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
    |||
Db 901 GACAAGTGCAGGTTCAAGCCCAAGCTCTGCGCGCATCCGCTGAGAGGCTCCAGGAAG 960
QY 961 gtggagagagaaagaccccgaggac 987
    |||
Db 961 GTGGAGAGGAAGACTTGCAGGGGAG 987

```

```

RESULT 7
LOCUS HS885L7 160241 bp DNA PRI 09-JUL-2000
DEFINITION Human DNA sequence from clone RP5-885L7 on chromosome
ACCESSION AL035669
VERSION AL035669.43 GI:8979786
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 160241)
AUTHORS Smith, M.
TITLE Direct Submission

```

JOURNAL COMMENT

Submitted (09-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Jul 8, 2000 this sequence version replaced gi:8919619.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW, SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr20>
 RP5-885L7 is from the library RCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP5-885L7 The true left end of clone RP4-563E14 is at 140192 in this sequence. The true right end of clone RP11-93B14 is at 44512 in this sequence.

FEATURES

source

```

1..160241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.2-13.33"
/clone="RP5-885L7"
/clone_1db="RCI-5"
2..64
/note="LIM49 repeat: matches 6210..6254 of consensus"
65..353
/note="AluJo repeat: matches 1..287 of consensus"
354..397
/note="LIM49 repeat: matches 6234..6302 of consensus"
4198..5157
/note="12 copies 80 mer 61% conserved"
complement(4223..5153)
/note="Tandem repeat. region contains two forced joins and single clone. Assembly consistent with restriction digest."
4229..5114
/note="443 copies 2 mer cc 59% conserved"
4231..5166
/note="234 copies 4 mer tccc 60% conserved"
4240..5151
/note="19 copies 48 mer 61% conserved"
4248..5159
/note="38 copies 24 mer 61% conserved"
4264..5163
/note="15 copies 60 mer 61% conserved"
4286..5152
/note="17 copies 51 mer 59% conserved"
4308..5171
/note="27 copies 32 mer 61% conserved"
4317..5156
/note="21 copies 40 mer 62% conserved"
4344..5142
/note="17 copies 47 mer 59% conserved"
4348..5145
/note="42 copies 19 mer 60% conserved"
4431..5144

```

repeat_region	/note="71 copies 34 mer 60% conserved"
repeat_region	4456. .5148
repeat_region	/note="71 copies 33 mer 60% conserved"
repeat_region	4486. .5151
repeat_region	/note="9 copies 74 mer 69% conserved"
repeat_region	4530. .5087
repeat_region	/note="18 copies 31 mer 71% conserved"
repeat_region	4533. .4822
repeat_region	/note="5 copies 58 mer 77% conserved"
repeat_region	4720. .5137
repeat_region	/note="11 copies 38 mer 61% conserved"
repeat_region	4833. .5064
repeat_region	/note="4 copies 59 mer 79% conserved"
repeat_region	5051. .5164
repeat_region	/note="19 copies 6 mer tctctc 63% conserved"
repeat_region	5934. .6864
repeat_region	/note="11M4 repeat: matches 4509. .5488 of consensus
repeat_region	6974. 7049
repeat_region	/note="4 copies 19 mer 80% conserved"
repeat_region	7224. .7451
repeat_region	/note="6 copies 38 mer 98% conserved"
repeat_region	7892. .8074
repeat_region	/note="MER68A repeat: matches 68. .210 of consensus"
repeat_region	8674. .8966
repeat_region	/note="AluI repeat: matches 5. .296 of consensus"
repeat_region	10876. .10962
repeat_region	/note="MIR repeat: matches 86. .162 of consensus"
repeat_region	11974. .12025
repeat_region	/note="13 copies 4 mer ctca 73% conserved"
repeat_region	14394. .14478
repeat_region	/note="5 copies 17 mer 77% conserved"
repeat_region	15824. .15891
repeat_region	/note="17 copies 4 mer tccc 67% conserved"
repeat_region	16266. .16577
repeat_region	/note="AluSg repeat: matches 1. .313 of consensus"
repeat_region	16578. .16666
repeat_region	/note="17 copies 17 mer 65% conserved"
repeat_region	16581. .16844
repeat_region	/note="11 copies 24 mer 68% conserved"
repeat_region	16582. .16891
repeat_region	/note="10 copies 31 mer 66% conserved"
repeat_region	16583. .16879
repeat_region	/note="9 copies 33 mer 66% conserved"
repeat_region	16585. .16840
repeat_region	/note="16 copies 16 mer 66% conserved"
repeat_region	16586. .16875
repeat_region	/note="5 copies 58 mer 72% conserved"
repeat_region	16590. .16874
repeat_region	/note="15 copies 19 mer 64% conserved"
repeat_region	16593. .16872
repeat_region	/note="7 copies 40 mer 67% conserved"
repeat_region	16598. .16837
repeat_region	/note="5 copies 48 mer 73% conserved"
repeat_region	16603. .16842
repeat_region	/note="3 copies 80 mer 77% conserved"
repeat_region	16638. .16901
repeat_region	/note="44 copies 6 mer aaagaa 62% conserved"
repeat_region	16664. .16885
repeat_region	/note="3 copies 74 mer 78% conserved"
repeat_region	16670. .16504
repeat_region	/note="5 copies 47 mer 77% conserved"
repeat_region	16848. .16899
repeat_region	/note="13 copies 4 mer gaag 100% conserved"
repeat_region	17287. .17411
repeat_region	/note="MER8A repeat: matches 35. .165 of consensus"
repeat_region	17447. .17486
repeat_region	/note="10 copies 4 mer ccat 92% conserved"
repeat_region	18303. .18302
repeat_region	/note="5 copies 40 mer 66% conserved"
repeat_region	18307. .18498
repeat_region	/note="4 copies 48 mer 67% conserved"
repeat_region	18317. .18496
repeat_region	/note="3 copies 60 mer 74% conserved"

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
Repeat_Region	18319. .18486	/note="7 copies 24 mer 66% conserved"				
Repeat_Region	18320. .18495	/note="44 copies 4 mer acac 64% conserved"				
Repeat_Region	18321. .18496	/note="11 copies 16 mer 64% conserved"				
Repeat_Region	18325. .18486	/note="9 copies 18 mer 66% conserved"				
Repeat_Region	18327. .18486	/note="8 copies 20 mer 66% conserved"				
Repeat_Region	18497. .18546	/note="25 copies 2 mer ac 72% conserved"				
Repeat_Region	19329. .19433	/note="LTR16C repeat: matches 286. .387 of consensus"				
Repeat_Region	19455. .19590	/note="4 copies 34 mer 88% conserved"				
Repeat_Region	19457. .19604	/note="2 copies 74 mer 81% conserved"				
Repeat_Region	19546. .19625	/note="4 copies 20 mer 76% conserved"				
Repeat_Region	19839. .19954	/note="2 copies 58 mer 87% conserved"				
Repeat_Region	20020. .20103	/note="MIR repeat: matches 51. .137 of consensus"				
Repeat_Region	20704. .20805	/note="3 copies 34 mer 90% conserved"				
Repeat_Region	20710. .20811	/note="2 copies 51 mer 90% conserved"				
Repeat_Region	20722. .20806	/note="5 copies 17 mer 89% conserved"				
Repeat_Region	21349. .21410	/note="2 copies 31 mer 93% conserved"				
Repeat_Region	21584. .21654	/note="LTR16A repeat: matches 88. .158 of consensus"				
Repeat_Region	21779. .21852	/note="Alu repeat: matches 238. .311 of consensus"				
Repeat_Region	24787. .24847	/note="MER1B repeat: matches 1. .63 of consensus"				
Repeat_Region	24855. .25064	/note="MIR1D repeat: matches 107. .348 of consensus"				
Repeat_Region	25853. .26065					
Query Match	43.7%; Score 699.8; DB 35; Length 160241;					
Best Local Similarity	95.7%; Pred. No. 9e-96;					
Matches 719; Conservative	0; Mismatches 32; Indels 0; Gaps 0;					
QY 633	ccagaacctgaactggtgcagcaccacacacacctccgcatcacagcatcctcaagtgc	692				
DB 61392	CCGGATCTCTCTCCAGCAGCCGACGACACACACCTTCGATCACACGACATCTCACTGCTCCT	61451				
QY 693	gtgtgaagctgaagccctcgaagcacttcacagcgccacatgtgcgtctcttcctcgaagaagac	752				
DB 61452	GGGTAGCTGGGCGCTTCGAGACATTCACAGCGCGCGCTGTGCGCTTCTTCTCTGGAAGAGAC	61511				
QY 753	gctgtgtgcggcggaagctcgcgcgggggtgcgcgcagagtgacctcgtgactactatgttcgc	812				
DB 61512	GCTGTGGGGGGGAGAGTCGCGGGGGGTGCGGAGAGTCCTCGAGACTACTTCATGTTTCG	61571				
QY 813	cgtgcgtgtccgacacacacagcgccgcacagtgtgtcacttcgctcgtgggaacattccggcc	872				
DB 61572	CGTGGCTGCGCGACACACGCGCGCGACACTGGTGCATCTTCCTCGGGAGCATTCGCCGCC	61631				
QY 873	ccagtcgaagttctgctctcgtggggcccccaagaagctgtggaggtgtcaagccagctcct	932				
DB 61632	CCGCTGCAGATTCTGTTCTGGGGCCCCCAAGACACAGCTGGGAGAGTTCAAGCCCAAGCTTCT	61691				
QY 933	gccgcatccgctcgtcgaaggtccagaagaagtgtgaaggaagaagcccccggaagaccgca	992				
DB 61692	GCCCATCGCGGTGCGAGGGCTCCAGGAAGTGAAGAGGAAGGAAGCCCGGGGAGCCCGGA	61751				
QY 993	ccaagagccagacacacagaggtcggactgtgtggccagagcatagcaaggttggggcag	1052				
DB 61752	CCACAGGCGACAGCACCCAGGCTGCGACTGTGGGCGCAAGAGCATAGCAAGGGGTGGGGCGAG	61811				

QY	1053	ggtgtgacgaggggccccacgcgagcgtgtgagccccagatgtggaccctgtgagag	1112
Db	61812	GGTGTGACGAGGGGCCCCACGACGAGCGTGTGAGCCCCACGATGCGGACCCCTGTGGAGAG	61871
QY	1113	gaaccagggggtatgaagcaggggcccacggggaagatagcggagcccttaagcccca	1172
Db	61872	GAGCCAGGGGGAGTAGAGCGAGGGGCCACGGGGAAAGTAGCCCGAGCCCTTAAGCCCCAA	61931
QY	1173	agagagcaagaaagaggaacctgtgagcttgagcccgcgaggacgcgccacagagccag	1232
Db	61932	AGAGAGCAAGAAGAGAGAACTGTGAGCTGAGCCGGCGAGACAGCCGCCACAGAGCCAGG	61991
QY	1233	ccctcaggaatgctcctcaggaagtgtgagaagaatgcgtctgaatttgaggagggtgtccctcag	1292
Db	61992	CCCTCAGATGCTCTCAGAGGTGAGAAAGATCGCTCGAATTTTGAGAGGGGTGTCCTCTCAG	62051
QY	1293	ccagagcagcctcagacagcggggacccaggaagtgtggcgctcagagaccttgaggagccctc	1352
Db	62052	CCAGGCGACCTCTCAGGACCGGGGACCCAGAGATGTGGGGCGGTCTCAGGACCTCTGGGAGGCACT	62111
QY	1353	ctgtccctgtctgcaagggggcttgggacctcg	1383
Db	62112	GCAGCCCTGCTCCGACCAACCTCTGGAGGCCAGG	62142
RESULT	8		
LOCUS	AF156878	2208 bp	16 th JAN-2000
DEFINITION	Rattus norvegicus oploid growth factor receptor mRNA, complete cds		
ACCESSION	AF156878		
VERSION	AF156878.1	GI:6693831	
KEYWORDS			
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 2208)		
AUTHORS	Zagon,I.S., Verderrame,M.F., Allen,S.S. and McLaughlin,P.J.		
TITLE	Cloning, sequencing, expression and function of a cDNA encoding a receptor for the oploid growth factor, [Met(5)]enkephalin		
JOURNAL	Brain Res. 849 (1-2), 147-154 (1999)		
REFERENCE	2 (bases 1 to 2208)		
AUTHORS	Zagon,I.S., Verderrame,M.F., Allen,S.S. and McLaughlin,P.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUN-1999) Neuroscience and Anatomy, Penn State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA		
FEATURES			
SOURCE	Location/Qualifiers		
	1..2208		
	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/tissue-type="brain"		
	/dev_stage="18 day-old fetus"		
	151..1893		
	/note="[mel-5]-enkephalin binding protein"		
	/codon_start=1		
	/product="oploid growth factor receptor"		
	/protein_id="AAF25002.1"		
	/db_xref="GI:6693832"		
	/translation="MDPDDCSTHEESEEDGEGQADDTTDEDTGDDGDAEARPSES LFQSRMGTRMRWAMQDMQRTNRHNPDTLDDCDGMDKCNLSFYNETICFPNGALIEE ILQWKNQNYIDLLENHSHYIOWLEPLREBQVWMAKPLTLKVEFKSKYEVRRLIVRA YELMLRGFGEHLEDRTGAVCARQNPQPRFHNLSHSHNNLRILKSLGLELEHFW QAPLYRFELEETLQVHKLPYVRGALDYFLFAVRCRQRELIVYFAMEHFKPREFEW GPRKLRKRPQTIPOPLTGGADKDGSDRPSQAGTGTGRTGSGRDI.SGDSGTAA DPSLINTKPSNGGTLGDNORPEAKSIPKSKRKKLGSNREQVPGENDPGQVSEVER IALNLECAASPIQDERREAPPPVAVRAVEAKRRKRVESGAGDGVSTQWQABA LQVPEPSPEAKQDNGPEDSNSQVGAEDSKSQVGPEDPNSQVLEDPNSQVGPEDPN SQVPEPSNPNSQVGPEDPNSQVGPEDPNSQVGPEDPNSQVGPEDPNSQVGPEDPN		

repeat_region	1549..1764	BLART5AEPKP"
	/note="imperfect"	
	/rpt_type=direct	
	/rpt_unit=1549..1575	
polya_signal	2031..2036	
polya_site	2203	
BASE COUNT	522 a 615 c 675 g 396 t	
ORIGIN		
Query Match	43.5%;	Score 696.6; DB 12; Length 2208;
Best Local Similarity	72.3%;	Pred. No.1.7e-94;
Matches 944;	Conservative 0;	Mismatches 334; Indels 27; Gaps 2
QY	48	ctgcgaactccacctggagaggagacgagagatgcggaagcgcggagagacgagactg 107
DB	135	CTCTCTGCTTGTGAGATATGAGACACCGGACCTGCATTTCCACTTGAGAGAGAGAGCGCA 194
QY	108	ctgaagacggcgagggccggccggcgcgagaggagcggagacggagagagagagagc 167
DB	195	GGAGATGCGCGAGATGCGCAGGCGGAGATATACGACCGATGAGACACGGCGAGCATGA 254
QY	168	ggaagagcccgcgcgcgcgccagctcgtcttcagttccagaatcaagaagaggtccagaa 227
DB	255	CGGCGACGCGGAGAGAGCGCGCCAAAGCTGTTCCAGTCCAGATGACAGGGTTACGGAA 314
QY	228	ctggcgagagcccaagaggagatgtgtagatattcgcaactatccgatatctgttggagc 287
DB	315	CTGGGCTGCTATGACAGGACATCAAAAGATACCGGACAACTACCCGGATTTGACAGATCA 374
QY	288	agactgcgaattggggacacgcgcaaacctagttcttcacagaatagatccgtctcgtcc 347
DB	375	AGACTGCATGCTGACATGCTGCACTGACACTGAGCTTTACAAAAATGAGATCTGCTTCCAGCC 434
QY	348	caagcgctgttcatattgaggagacattcttcagaactcgagcagcaactatgaccttcctga 407
DB	435	AAATGGGGCTCTCATCGAGGACATTTCTTCAGAACTGGAAGAACACTATGACTCTCTCGA 494
QY	408	ggacatacactctcatatccagctgtctgttccctcttgtagagacagagatgaacttgca 467
DB	495	AGAGAAATCACTCTCATATCCAGTGGCTGTTCTCTCGCGGAAACAGAGATGAATCTTGCA 554
QY	468	tgcgaagccctcaccgctcagggagggtcgaggtgtttaaagcttccagagatccagaaga 527
DB	555	CGCCAAAGCCCTCAACCCCTGAAGAGAGGTGTGAGGCATTTAAAAGCTCAAGAAATCAGAGA 614
QY	528	ggcggtttccgggggcttcagagctatctatcttgggtcttcaaggatccggctcgagagacc 587
DB	615	GGGTTTGTTCGGGGCTTATGAGCTATCTGTGGCTTCTATGAGGTTCACCTTGAAGACCG 674
QY	588	aggcagcgggcagcgtgtggtccagacacagaactaccagaagcgtctccagaaccttaactg 647
DB	675	GGGCGAGGGGTGCTGATGACCGGTGCAAGAACTTCAAGCCGGGGCTTCCAAATCTGAAAG 734
QY	648	ggcgagggccaacaacacctccgcatlaccagcatctcaagtgcgcgtgtgagtgagcct 707
DB	735	CCACGCGCAACAACACTGCGATTATACAGCATCTCAAGTCACTGGGTGAGCTGGGCTT 794
QY	708	cgaggaacttcaaggcgcaactggttcggtttcttccttgtagagagcgtcgtggtcgcgga 767
DB	795	AGAAACTACACAGGAGACCCCTGTCGCTTTCTCTGAGAGAGACCTTGTATCAGACACA 854
QY	768	gctgcgggggtgtcggaagatgcctctgactacttaattgttcgcgtgtgcgttcgcgaca 827
DB	855	ACTGCGACGCTGCGCCAGAGATGCCCTGAGACTACTTCTGTTCCTGTGCGCTCGCCCGGCA 914
QY	828	ccaagcgccggcagcgtgtggaacttcgccttgtagaaccttcggcccgcgtcgaagtgcgt 887
DB	915	CCAAGCGCCGGAGCTTGTGTACTTTGCTTGGAAGACTTCAAGCCTCGCCAGAGATTGT 974
QY	888	ctgggggcccacaagaacagctctgcgagggtttaaagcccagctctctgcgcgcatccgctcga 947

Dh 975 CTGGGGGCCCCGTGACCAAGCTGCGAGATTCAGGCCCAACATACCCCAAGCCACTGAC 1034

Qy 948 gggtcccaagaggttgagggaggaagagcccccggagcccccagccagagccagac 1007

Dh 1035 GGGACCAAGGAGGAGATTAAGATGAGGGCTCCAGGAGCCCTCCCAAGAGGCTGGGAC 1094

Qy 1008 ccaaggtcgcagccttgaggccagagcatagcaaggttgaggagagag--tgagagag 1064

Dh 1095 CCAGGCTGGAACCTGTGATCTGGAAGGAGCTGAGTGGGAGCACTGGAACAGCTGAGGA 1154

Qy 1065 gcccaagccagagagcgttgagagcccccagagtcggagagcccttgagagagagagagga 1124

Dh 1155 TCCCTCACTGCTGAACAACAAGCCCTCAGATGAGGGGAGCACTTGATGGAACCAAGAGGA 1214

Qy 1125 tgaagcagaggggcccagggagagatagagccggagcccttaagcccccaagagagagaa 1184

Dh 1215 TGAAGCTA-----AGTCCCTGAGTCCCAAGAGAGCAAGAA 1250

Qy 1185 gaggagcctgagctgagccgagccgagagcagcccccagagccagagccctcagagtc 1244

Dh 1251 AAGGAGTGGAGGAGGAGCAAGGAGCAAGGAGCTCCAGGAGGAGCAATCCCAAGGCTGT 1310

Qy 1245 ctcaagagtgagagagatcgctctgaatttgagaggggtgagccctcagccagagagcct 1304

Dh 1311 CTCTGAGGTAGAGAAATGCTTACCTTGAAGAGTGTGCGCTTACCTTACCTATCAGCA 1370

Qy 1305 cagagcagggagccagagagtgaggggtcagagccctggagagc 1349

Dh 1371 GGAGCCCAAGGAGGCTGAACCGCTGTCTGTGGCCAGGAGGTC 1415

RESULT 9

AF112980 6670 bp DNA PRI 10-JAN-1999

LOCUS AF112980 Homo sapiens 7-60 gene, complete cds.

DEFINITION AF112980

ACCESSION AF112980

VERSION AF112980.1 GI:4139271

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

EXTRACT: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 6670)

AUTHORS Takanosu,M., Itoh,J. and Mayne,R.

JOURNAL Genomic Structure of Human Gene 7-60

TITLE Unpublished

REFERENCE 2 (bases 1 to 6670)

AUTHORS Takanosu,M., Itoh,J. and Mayne,R.

JOURNAL Direct Submission

REFERENCE Submitted (10-DEC-1998) Cell Biology, University of Alabama at Birmingham, VH 302, Room 605, Birmingham, Alabama 35294, USA

JOURNAL Location/Qualifiers

FEATURES

source

1..6670

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q13.3"

join(<206..268,906..984,2235..2313,3147..3213,4125..4274,4891..>6309)

/product="7-60"

join(206..268,906..984,2235..2313,3147..3213,4125..4274,4891..6309)

/codon_start=1

/product="7-60"

/protein_id="A03745.1"

/db_xref="GI:4139272"

/translation="MTGSRMRATRDRCRYRHNYPDLVERDCNGDFNLSFYRNEIRF LPNCGFTEDIQNTDNYDLLEDNHSYICRMFIREGVNMAHAPLILREYEVKSSQ EIOERLRYATNAHAGLDRPAGRGHGGSTLEPDLPELRSNNKNTITLKS LGELGLHFQAPLVRFLLESILVRLEPVQASLDYFMFVAVGCRHGRDLVFAWH FRPCKFLHGFQAPLVRFLLESILVRLEPVQASLDYFMFVAVGCRHGRDLVFAWH KGGRVDEGPQPRSEVPEODAGPLERSGDEGHEGDEPELSPKESKKRLELSRE QPTEPEQPQASSEVERIALNLGSCALSGSLRTGTGAGGDDPEANVPORPGRARV

ADVRKRRKVDGAGDAAVAASGAQTLALAGSPAPSGHPRKASHSENGEEDTEGRTG
PKETGSPSEETGSPAPAGDEPAESPSEETGSPAPAGDEPAESPSEETGSPAP
GAPADPRAPKPSPTSPSPAPGPRDRPAPSPSPSPSPAPGAPDPRAPSPSPSP
PAPADDEPSPSPSPSPAPGPRDRPAPSPSPSPAPGAPDPRAPSPSPSPSPSP
BASE COUNT 1367 a 1998 c 2169 g 1136 t

ORIGIN

Query Match 43.3%; Score 693.4; DB 11; Length 6670;

Best Local Similarity 95.2%; Pred. No. 3.1e-94;

Matches 715; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 633 ccaagacctgagctgagggagagcccaacacccctcgatcacacgcatcctcaagtgc 692

Dh 4875 CCCGATCTCTCGAGGCGACCCCAACACCTTCCATCACGCACTCTCAAGTGCT 4934

Qy 693 gttgagctgagccttcagacatctccagagcgccatggtcgccttcttcctgagagac 752

Dh 4935 GGGTAGAGCTGGGCTCCAGAGCACTTCCAGGCGCCGCTGTCGCTTCTTCTGGAGAG 4994

Qy 753 gctgtgagggagagctgagcggaggtgagagagagagagagagagagagagagagag 812

Dh 4995 CTTGATGCGGCGGAGCGCTCCGCGGGGTGCGGCAAGTGCCTGACATTCATGTTCCG 5054

Qy 813 cgtgagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 872

Dh 5055 CGTGGCTGCGGAGACCAACGCGCCCAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 5114

Qy 873 ccgctgcaagtcgtctggtgggcccccaagaagctgagaggttcaagccagctctct 932

Dh 5115 CCGCTGCAAGTTCGTCTGGGGGGCCCCCAAGAACCTCGGAGGTTCAAGCCACCTCTCT 5174

Qy 933 gccgcatcgcctgagaggtcctcaagaggtgagagagagagagagagagagagagagag 992

Dh 5175 GCCCATCCCGCTCCAGGAGCTCCAGGAAGGTGAGAGAGAGAGAGAGAGAGAGAGAG 5234

Qy 993 ccagagagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1052

Dh 5235 CCAGGAGGCGACGCCACCAAGGCTCGGACTGTGGGCCAGAGCATAGCAAGGTTGGGGCAG 5284

Qy 1053 ggtgagagaggggcccccaagcagcagcagcagcagcagcagcagcagcagcagcagcag 1112

Dh 5295 GGTGAGAGAGGGGCCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5354

Qy 1113 gagccaggggagatgagcaggggggggggggggggggggggggggggggggggggggg 1172

Dh 5355 GAGCAGGAGGAGATGAGCAGGAGGAGCAGGAGGAGAGATAGGCGGAGCCCTTAACCCCAA 5414

Qy 1173 agagagcaagagagagagagagcctgagcagcagcagcagcagcagcagcagcagcagcag 1232

Dh 5415 AGAGAGCAAGAGAGAGAGAGCTGGAGCTGAGCCGGGAGAGAGAGAGAGAGAGAGAGAG 5474

Qy 1233 cctcagagtgctcctcaagagtgagagagatcgctctgaatttgagaggggtgagcctcag 1292

Dh 5475 CCCACAGAGTCTCCACAGAGGTGAGAGAGATGCTGATTTGAGAGGGGTGTGCCCTCAG 5534

Qy 1293 ccaagggcagcctcagagcagggagagcccaagagtgagcagcagcagcagcagcagc 1352

Dh 5535 CCAGGAGAGCCTTCAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5594

Qy 1353 ctgtcctgtcgtcagggggtcgtgagcctcag 1383

Dh 5595 GCAAGCCCTGCGCCCAACCCCTGGAGGAGGAG 5625

RESULT 10

I66494/c 166494 bp DNA PAT 28-DEC-1997

LOCUS I66494 Sequence 14 from patent US 5670367.

DEFINITION I66494

ACCESSION I66494

VERSION I66494.1 GI:2724471

KEYWORDS Unknown.

[illegible]


```

----- Project Information
Center project name: L5033
Center clone name: 178_A.24
----- Summary Statistics
Sequencing Vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 202880 bases at least Q40
Consensus quality: 208799 bases at least Q30
Consensus quality: 210932 bases at least Q20
Insert size: 215000; agarose-ff
Insert size: 212021; sum-of-ctnigs
Quality coverage: 4.8 in Q20 bases; agarose-ff
Quality coverage: 4.8 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1080: contig of 1080 bp in length
* 1081 1180: gap of 100 bp
* 1181 3071: contig of 1891 bp in length
* 3072 3171: gap of 100 bp
* 3172 4943: contig of 1772 bp in length
* 4944 5043: gap of 100 bp
* 5044 6632: contig of 1589 bp in length
* 6633 6732: gap of 100 bp
* 6733 8188: contig of 1456 bp in length
* 8189 8288: gap of 100 bp
* 8289 10317: contig of 2029 bp in length
* 10318 10417: gap of 100 bp
* 10418 11053: contig of 636 bp in length
* 11054 11153: gap of 100 bp
* 11154 13839: contig of 2686 bp in length
* 13840 13939: gap of 100 bp
* 13940 17082: contig of 3143 bp in length
* 17083 17182: gap of 100 bp
* 17183 20437: contig of 3255 bp in length
* 20438 20537: gap of 100 bp
* 20538 24737: contig of 4200 bp in length
* 24738 24837: gap of 100 bp
* 24838 28414: contig of 3577 bp in length
* 28415 28514: gap of 100 bp
* 28515 32961: contig of 4447 bp in length
* 32962 33061: gap of 100 bp
* 33062 40085: contig of 7024 bp in length
* 40086 40185: gap of 100 bp
* 40186 46580: contig of 6395 bp in length
* 46581 46680: gap of 100 bp
* 46681 54064: contig of 7384 bp in length
* 54065 54164: gap of 100 bp
* 54165 63367: contig of 9203 bp in length
* 63368 63467: gap of 100 bp
* 63468 73091: contig of 9624 bp in length
* 73092 73191: gap of 100 bp
* 73192 82289: contig of 9098 bp in length
* 82290 82389: gap of 100 bp
* 82390 91847: contig of 9458 bp in length
* 91848 91947: gap of 100 bp
* 91948 103913: contig of 11966 bp in length
* 103914 104013: gap of 100 bp
* 104014 115132: contig of 11119 bp in length
* 115133 115232: gap of 100 bp
* 115233 126702: contig of 11470 bp in length
* 126703 126802: gap of 100 bp
* 126803 138890: contig of 10088 bp in length
* 138891 139990: gap of 100 bp
* 139991 153374: contig of 13384 bp in length
* 153375 153474: gap of 100 bp

```

```

FEATURES
source
* 153475 168194: contig of 14720 bp in length
* 168195 168294: gap of 100 bp
* 168295 182165: contig of 13871 bp in length
* 182166 182265: gap of 100 bp
* 182266 214721: contig of 32456 bp in length.
Location/Qualifiers
1. .214721
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-178A24"
/clone_id="RP23-178A24"
1. .1080
/note="assembly_fragment"
1181. .3071
/note="assembly_fragment"
3172. .4943
/note="assembly_fragment"
5044. .6632
/note="assembly_fragment"
6733. .8188
/note="assembly_fragment"
8289. .10317
/note="assembly_fragment"
10418. .11053
/note="assembly_fragment"
13940. .17082
/note="assembly_fragment"
17183. .20437
/note="assembly_fragment"
20538. .24737
vector_side:right"
vector_end:right"
24838. .28414
/note="assembly_fragment"
28515. .32961
/note="assembly_fragment"
33062. .40085
/note="assembly_fragment"
40186. .46580
/note="assembly_fragment"
46681. .54064
/note="assembly_fragment"
54165. .63367
/note="assembly_fragment"
63468. .73091
/note="assembly_fragment"
73192. .82289
/note="assembly_fragment"
82390. .91847
/note="assembly_fragment"
91948. .103913
/note="assembly_fragment"
104014. .115132
/note="assembly_fragment"
115233. .126702
/note="assembly_fragment"
126803. .139890
/note="assembly_fragment"
139991. .153374
/note="assembly_fragment"
153475. .168194
/note="assembly_fragment"
168295. .182165
/note="assembly_fragment"
182266. .214721
/note="assembly_fragment"
BASE COUNT 57421 a 48869 c 48673 g 57057 t 2701 others
ORIGIN

```



```
/protein_id="BAA01269.1"
/db_xref="GI:221788"
/translation="MATDIDMLDGLDLSSELEEDALERDERGRDDPESDSSGSC
SSDDEDEWDEPGDGAELIDAIKPGPARPEDGTPEASTPRAARADDEPPAT
GWSRLGTNRASAPREPHRGKVARIQPPTAPPRKGRGRGRGRGRGRGRGR
KPRRVRNMANOCGRHPASARIDPGATHEGARGEOQLDSGGPRGRGRQAPPL
MASLTPPHADGAPVERKAPASADTIDPAVRAVLRISERAAVERISESFRALYV
ODPEGMPFPAANSPPMAPVLATAGSGDEARRRSMETLVAGHSPLRTAANRAAS
TAKAMRCVLRQENLITALLASADETLAKMCKIHNNPIPIRQDPICTAAVLENLAT
RVPFLQCYLAKRGLGDLDCSRRRLSDIKDINSFVLITARIANRVEGVSDIDT
TVVGAGEITHMFYIPGACMAGLIEILDTHROECSSRVCETASHTIAPLYHGKXYFC
NSLP"
2502..2507
/feature="for UL54"
2711..3271
/codon_start=1
/product="UL55 ORF"
/protein_id="BAA01270.1"
/db_xref="GI:221789"
/translation="MTTTPLSNLFBRADITHVAPPYCINATQNAENALHTKTDPAK
LAARSYVARSCTSGPIHCFPFVYVYDSSHSLPLVYELNRADLVNHPYVLELXK
RGRRLCTGPPSCGTLKDVSGASPAGEITNGIVYHCHCRIPESKTCWLASALQHL
RSISSGTAARAAEQRRHKIKIKIV"
3310..3315
/feature="for UL55"
complement(3413..3418)
/feature="for UL56"
complement(3444..4151)
/codon_start=1
/product="UL56 ORF"
/protein_id="BAA01271.1"
/db_xref="GI:221790"
/translation="MALGASHAHACRDGDDSDVIDAPPYEVSAAGAGQFVIDIT
PTSPPTYSAGTSPVGLVSPASGSDVECEGRKRAKRAKRAKRAKRAKRAK
GPGGLEVEPLFLPETMIGAPGVGDLPGLPYATATSDRPTVYMAACPTERP
GGGVGADQDQVYOSSTRTMRPPLVNSRELRYAORAKARCASSDTPQAGMGCGICRHAY
FGVAVVYVVIILAFIMR"
4355
/feature="right extremity of long unique region"
4356
/feature="start of internal large repeat"
6594..8808
/feature="latency associated transcript (pct.)"
complement(7996..8001)
/feature="for RL2"
complement(join(8251..9974,10156..10834,11242..11316))
/feature="RL2 ORF"
/feature="Vmw118 (ICP0)"
/codon_start=1
/product="Vmw118 (ICP0)"
/db_xref="GI:2626942"
/translation="MEPRPGTSSRADPGEPRPPROTGTQPAAPHANGMLNDQMVLAS
SDSEETEVG1ISDDDLHRDSTSENGSTDETFEGLMDAATPPARPAEROGSTPAD
AOGSCGGGPGVEEAEAGGGDCAVCTDEIAPLRQSPCLPFCIPCKMTIPIR
NTPLCTGTPAVYLVGTAGSGSFSTIPIVMDPRTRVAAVRAAGTAVDPMTGNGPT
APRSLSGHTVRLSPTPMPCGDDDDDLADVDYVPAAPRAAPRGGGAGCATRGT
SQPARTRPAPGAPRRSSSGGAPLRATVSGSGGGGPRVAAVPRVATLPPAAGGRGQ
ARRVEDAAAEAGKTPPARQPRAPQEPPIVISDPPSPRRAPGGLSVSSSAV
SSGPGGGGLPQSSGRAPRAVAPRAVRSPPRAAAVSAADAAAPAPAPAVVDH
RABPRMTQATQDTQOASLGRAGTIDARGSGGPGAEPPGPRNTNPGAPHAEEA
AABPRKRSGDSGPAASSASSAAPPSPPLAPOGVGAARAPAPSDSGDRHGFL
APASGAAPSPASPSQAAYAAASSASSASSASSASSASSASSASSASSASSA
SSSAGGAGGVSASGGERETSLGRAPAPPRKCAKTRIAEGEPGAPADPPAP
GLTRILPIAGVSSVALAPYVNTKVTDCPLVDMETGHGAYVVLVDQGVNADLLR
AAAPAMSRRTLLPEHARNCVPRPDYPTPPAPSEWNSLMWTPVGNMLFDQGTGLVGLDHF
GLRSRHPMSREOGAPAPADAGHGE"
complement(<8251..9974)
/number=3
/product="Vmw118 (ICP0)"
complement(10156..10834)
/number=2
/product="Vmw118 (ICP0)"
complement(<11242..11316)
```

```
/number=1
/product="Vmw118 (ICP0)"
complement(11474..11478)
/feature="for RL2"
complement(11876..11881)
/feature="for RL1"
complement(join(12240..12530,12685..13179))
/feature="RL1 ORF"
/codon_start=1
/product="neurovirulence factor (ICP34.5)"
/protein_id="BAA23428.1"
/db_xref="GI:2626943"
/translation="MSRRGRRRRRPPRRPGAPVAPRGAPVPPGALPTADSGM
VPAVDSTAVESAPASSLRLRLVLPQADSDPDADNDDEMANSPSEGGKAP
EAPHAAPAAAPPPPRKRGPPQPLPPLALURTTETYLRLSKRRPPAPSPAD
APRGKVCSPRVQVRLVAVETARLARSGSMARERDRFRFRRAAAAVICPCLE
PEARARARARAHEDGPPAEAEAAAGSAAAGPGRRAV"
complement(<12240..12530)
/number=2
/product="neurovirulence factor (ICP34.5)"
complement(12685..13179)
/number=1
/product="neurovirulence factor (ICP34.5)"
complement(13338..13341)
/feature="for RL1"
/feature="terminally redundant element(a sequence)"
13619
/misc_feature
/feature="start of c sequence"
BASE COUNT 1874 a 5124 c 4789 g 1893 t
ORIGIN 0.74-0.84 genome % units.
```

```
Query Match 4.18; Score 65.2; DB 73; Length 13680;
Best Local Similarity 62.0%; Pred. No. 0.29;
Matches 103; Conservative 0; Mismatches 63; Indels 0; Caps 0;

QY 27 gccagacatgagcagcccgagctcgactccactgtagagagagagagatcgga 86
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8809 GCCGACCGCGGCGCGAGCGCTCCACACGACCCCGCCAGAGAGCGGAGAGGA 8868

QY 87 ggaacgagagagcagagctcgagagcgcgcgcgcgcgcgcgcgcgcgcgc 146
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8869 GGAGCGGAGAGAGAGAGCGCGGAGAGAGAGCGGAGAGAGAGCGGAGAGGA 8928

QY 147 agggagcagagcagagagtcgagagcgcgcgcgcgcgcgcgcgcgcgc 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8929 GGAGCGGAGAGAGAGAGAGCGGAGAGAGAGAGCGGCGCGGCGAGCC 8974

RESULT 15
HSV2HG52 154746 bp DNA VRL 12-DEC-1999
DEFINITION Herpes simplex virus type 2 (strain Hc52), complete genome.
ACCESSION 286099
VERSION 286099.2 GI:6572414
capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyuridine triphosphatase; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; RL1 gene; RL2 gene; RSL gene; tegument protein; thymidine kinase; UL1 gene; UL10 gene; UL11 gene; UL12 gene; UL13 gene; UL4 gene; UL5 gene; UL6 gene; UL7 gene; UL8 gene; UL9 gene; UL20 gene; UL21 gene; UL22 gene; UL23 gene; UL24 gene; UL25 gene; UL26 gene; UL26.5 gene; UL27 gene; UL28 gene; UL29 gene; UL3 gene; UL30 gene; UL31 gene; UL32 gene; UL33 gene; UL34 gene; UL35 gene; UL36 gene; UL37 gene; UL38 gene; UL39 gene; UL4 gene; UL40 gene; UL41 gene; UL42 gene; UL43 gene; UL44 gene; UL45
```

SOURCE ORGANISM human herpesvirus 2.
 human herpesvirus 2.
 Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
 Alphaherpesvirinae; Simplexvirus.
 1 (bases 1 to 154746)
 McGeoch,D.J., Moss,H.W., McNab,D. and Frame,M.C.
 DNA sequence and genetic content of the HindIII 1 region in the
 short unique component of the herpes simplex virus type 2 genome:
 identification of the gene encoding glycoprotein g, and
 evolutionary comparisons
 J. Gen. Virol. 68 (Pt 1), 19-38 (1987)
 8711457
 2 (bases 1 to 154746)
 Everett,R.D. and Fenwick,M.L.
 Comparative DNA sequence analysis of the host shutoff genes of
 different strains of herpes simplex virus: type 2 strain HG52
 encodes a truncated UL41 product
 J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)
 90278430
 3 (bases 1 to 154746)
 McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.
 Comparative sequence analysis of the long repeat regions and
 adjoining parts of the long unique regions in the genomes of herpes
 simplex viruses types 1 and 2
 J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)
 92113549
 4 (bases 1 to 154746)
 Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J.
 A novel herpes simplex virus gene (UL49a) encodes a putative
 membrane protein with counterparts in other herpesviruses
 J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)
 92356101
 5 (bases 1 to 154746)
 Dolan,A.
 Direct Submission
 Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church
 Street, Glasgow, G11 5JR, UK
 revised by [6]
 6 (bases 1 to 154746)
 Dolan,A.
 Direct Submission
 Submitted (08-DEC-1999) A. Dolan, MRC Virology Unit, Church
 Street, Glasgow, G11 5JR, UK
 On Dec 13, 1999 this sequence version replaced gi:1869820.
 Location/Qualifiers
 1..154746
 /organism="human herpesvirus 2"
 /strain="HG52"
 /db_xref="taxon:10310"
 1..254
 /note="a sequence"
 255
 /note="start of TRL (terminal copy of Long Repeat region)."
 440..934
 /gene="RL1"
 /number=1
 440..1379
 /gene="RL1"
 join(440..934,1089..1379)
 /gene="RL1"
 /codon_start=1
 /protein_id="CAB06759.1"
 /db_xref="GI:1869821"
 /db_xref="SWISS-PROT:P28283"
 /translation="MSRRGRRRRGGRRRRPGAPVPRGAPVPPGALPTADSGM
 VPAYDSGTAIVSAPASSLKRWLLVPOADSDADYAGNDADAMANSPSEGGGKAP
 EAPHAAPAAACPPPPRRKGRPPRLPPLHALRLRTTYETLIARLSLRRRRPPASPAPD

APRKGCFSPRVQVRLVAMETARLARLRSGSWARERARDRFRFRVAAAENVIGPCLE
 PEARARARARARAEHDGPAEEEAASAAAGSSAAGCRRAY"
 446..460
 /rpt_family="7"
 461..547
 /rpt_family="6"
 943..1070
 /rpt_family="5"
 1089..1379
 /gene="RL1"
 /number=2
 /product="neurovirulence factor"
 1738..1743
 2303..5368
 /gene="RL2"
 2303..2377
 /gene="RL2"
 /number=1
 join(2303..2377,2785..3462,3644..5368)
 /gene="RL2"
 /codon_start=1
 /protein_id="CAB06760.1"
 /db_xref="GI:1869822"
 /db_xref="SWISS-PROT:P28284"
 /translation="MEPRGTSRADPGEPFRPQTPTQPAAPHAMGLINDQMGLAS
 SDSEETEVGISDDDLHRDSTSEAGSTDEMEFAGLMDATPPARPAPEROGSPTPAD
 AOSGCGGVPVEEAEAGGGDCAVCTEIEAPLRCQSPFCILPKPCIMKWIPLR
 NCPICNTPAVALIVGTAGSSTIPIVNDPTRYEAFAVAGVADPITMGNGRT
 APSLSIGHTVRLASPTPMPCTDDEDLDLAVDVPAPRAPRGAGGAGATGCT
 SOPAATMPAPPAAGPSSSGCARLRAGVSGSGGAPVAAYVRVASLPPAAGGGAQ
 ARVGEDEAAABEGRTPPAROPRAQGPPIYISDPPSPRRRPPGPISEVSSSAQV
 SSGGGGGQSSGKARAPRAVAPVRSPRRAPAAAPVVASADAGAPAPVAVDAH
 PAPSRRTQAQDTQOASLGRAGATPARSGGAGGVPVPGATPPAPAPAAAGCA
 AARPKRRGDSGPPAASSSAASAPRSPAPQGVGAKRPAAPRAPDSGDRGGLP
 APASAGAPPASPPSOAAVAASSSAASSSAASSSAASSSAASSSAASSSA
 SSSGAGGGSVAASGAGERRRTSLGPRAAPRGPKCAKRTRHAGGEPGAPADAP
 GLTRYLDINAGSVVALAPVYNTKYGDCCLPYIDMETGHGAYIVYDQTVADILR
 AAPAMSRRTLLPEHARKCYRPDPDPTPPASEMNSLMKTPVGNMFLDQGLVGLDFH
 GLRSRHPWSREGAPAPAGDAPAGHE"
 2785..3462
 /gene="RL2"
 /number=2
 3468..3607
 /rpt_family="4"
 3644..5368
 /gene="RL2"
 /number=3
 /function="modulator of cell state and gene expression"
 4655..4771
 /rpt_family="3"
 complement(4811)
 /note="proposed LAT splice acceptor site"
 5618..5623
 complement(7025)
 /note="proposed LAT splice donor site (5' end of stable
 LAT species)."
 complement(7732)
 /note="primary transcription initiation site for latency
 associated transcripts (LAT's)"
 9047..9193
 /rpt_family="2"
 9276..9297
 /rpt_family="1"
 9298
 /note="start of UL (Long Unique region)."
 9427..10101
 /gene="UL1"
 9427..10101
 /gene="UL1"
 /codon_start=1
 /product="viral glycoprotein L"
 /protein_id="CAB06761.1"
 /db_xref="GI:1869823".

```

/db_xref="SWISS-PROT:P28278"
/translation="MGFVCLFGLVMGAMGAMGSSQATEYVLSVIAKEVGDILRVPC
MRTPADVSWRVEAPSVIDYARIDGIFLRHCPLDIFLMDRHRORAYLVNPLEFAG
FLIEDLSHVSFPADTOETTRALYKEIRDALGSRKAVSHAPVAGCVNFDYSSTRRC
VGRDRLRPANTISTMEPPVSSDDEASSQSKPLATOPVYLAISNAPPRVSPTRGRRH
TRLRN"
gene
10211..10978
/contig="UL2"
/contig_start=10211
/contig_end=10978
CDS
10211..10978
/contig="UL2"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="CAB06762.1"
/db_xref="GI:1869824"
/db_xref="SWISS-PROT:P28275"
/translation="MFSSTTPPEQIGLSDGAPPLPTSVPLDMARFRAFLIDAR
PLEPELANPLTARLAEYDRKCTEYVLPREDVFSWIRYCTPDVYVYIGDPTF
HPOAHGLAFVRADVPVPPSLRNVLAVKNCYPDARMSRGCLKVARQGVLLNTT
LTVKGAASHSKLGMDFVGVVQRLAARRPGLVFMIMGAAHQAONAIRPDPQHYVLK
FSHSPSLSKYFEGFCQHFLLANRYLETREDIMPIDWSY"
11019..11024
11033..11734
/gene="UL3"
/codon_start=1
/protein_id="CAB06763.1"
/db_xref="GI:1869825"
/db_xref="SPTREMBL:Q89644"
/translation="MKSRSVSRVSMGVEERYPSAFTLLASGWTAFQNDHDPAS
PNTPIESIAGTADAHVGLDGEPPDRDAISPLTSVAGDPPGADGPFYVTEDLFWTS
SIDELGRDLTDTRKDLRLSLAKFSIACITSSFSGTAAORRGAPPORTCVPREN
KSLQMFVLCRRANAQVREQLRAVIRSRKPRKTYTRSDGRICPAVFVHEFVSSEP
polya_signal
gene
11033..11734
CDS
11033..11734
/gene="UL3"
/codon_start=1
/protein_id="CAB06763.1"
/db_xref="GI:1869825"
/db_xref="SPTREMBL:Q89644"
/translation="MKSRSVSRVSMGVEERYPSAFTLLASGWTAFQNDHDPAS
PNTPIESIAGTADAHVGLDGEPPDRDAISPLTSVAGDPPGADGPFYVTEDLFWTS
SIDELGRDLTDTRKDLRLSLAKFSIACITSSFSGTAAORRGAPPORTCVPREN
KSLQMFVLCRRANAQVREQLRAVIRSRKPRKTYTRSDGRICPAVFVHEFVSSEP

```

Query Match 4.18; Score 65.2; DB 73; Length 154746;

Best Local Similarity 62.0%; Pred. No. 0.1;

Matches 103; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

QY 27 gccgagcatgagagaccgactgctgactccacctggagagagagagatgcga 86
    ||| | ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 122440 GCCGGACCGGAGCGGATGCTCCACAGCCCGCCCGCAGAGAGAGCGGAGAGGA 122499

QY 87 ggaacgcgagagacgagactgcgagacgcgcgcgcgcgcgcgcgcgcgcgcgc 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 122500 GGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 122559

QY 147 agggagacgagagagagatcgagagagcgcgcgcgcgcgcgcgcgcgcgcgc 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 122560 GGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 122605

```

Search completed: October 12, 2000, 18:10:42
Job time: 35618 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 18:10:42 ; Search time 17897.3 Seconds
(without alignments)
572.995 Million cell updates/sec

Title: US-09-431-843b-9
Perfect score: 2348
Sequence: 1 tagaattcagcgccgctga.....cccttcgtgagtgagggggagcg 2348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_cm:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_pl3:*
34: gb_pl4:*
35: gb_pl5:*
36: gb_pl6:*
37: gb_pl7:*
38: gb_pl8:*
39: gb_pl9:*
40: gb_pl10:*
41: gb_pl11:*
42: gb_pl12:*
43: gb_pl13:*

44: gb_pl14:*
45: gb_pl15:*
46: gb_pl16:*
47: gb_pl17:*
48: gb_pl18:*
49: gb_pl19:*
50: gb_pl20:*
51: gb_pl21:*
52: gb_pl22:*
53: gb_pl23:*
54: gb_pl24:*
55: gb_pl25:*
56: gb_pl26:*
57: gb_pl27:*
58: gb_pl28:*
59: gb_pl29:*
60: gb_pl30:*
61: gb_pl31:*
62: gb_pl32:*
63: gb_pl33:*
64: gb_pl34:*
65: gb_pl35:*
66: gb_pl36:*
67: gb_pl37:*
68: gb_pl38:*
69: gb_pl39:*
70: gb_pl40:*
71: gb_pl41:*
72: gb_pl42:*
73: gb_pl43:*
74: gb_pl44:*
75: gb_pl45:*
76: gb_pl46:*
77: gb_pl47:*
78: gb_pl48:*
79: gb_pl49:*
80: gb_pl50:*
81: gb_pl51:*
82: gb_pl52:*
83: gb_pl53:*
84: gb_pl54:*
85: gb_pl55:*
86: gb_pl56:*
87: gb_pl57:*
88: gb_pl58:*
89: gb_pl59:*
90: gb_pl60:*
91: gb_pl61:*
92: gb_pl62:*
93: gb_pl63:*
94: gb_pl64:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2348	100.0	2423	11	AF172451	AF172451 Homo sapi
2	2263.6	96.4	2483	11	AF172453	AF172453 Homo sapi
3	2256.6	96.1	2423	11	AF109134	AF109134 Homo sapi
4	2212.8	94.2	2363	11	AF172452	AF172452 Homo sapi
5	1688.8	71.9	160241	35	HS885L7	AL035669 Human DNA
6	1680.8	71.6	6670	11	AF112980	AF112980 Homo sapi
7	1352.6	57.6	1676	11	AF172450	AF172450 Homo sapi
8	979	41.7	1232	11	AF172449	AF172449 Homo sapi
9	696.8	29.7	2208	12	AF156878	AF156878 Rattus no
10	122	5.2	162114	72	AF157706	AF157706 Human her
11	121.8	5.2	30752	72	HHU92288	U92288 Human herpe
12	121.6	5.2	161573	72	AB021506	AB021506 Human her

13	110.6	4.7	4742	72	AE015297	Human	herp
14	110.6	4.7	13965	72	AE015298	Human	herp
15	97.8	4.2	21191	72	HHU13194	Human	herp
16	94.6	4.0	159321	72	HHV6AGNA	Human	herp
17	89.4	3.8	167830	43	AC020678	Human	sapl
18	88.4	3.8	6192	5	AE030811	Sequence	
19	88.4	3.8	6449	5	AE030812	Sequence	
20	88.2	3.8	329621	47	AC062685	Human	sapl
21	86.6	3.7	192453	45	AC033920	Human	sapl
22	84.2	3.6	11177	32	PCBPPT	Human	sapl
23	81	3.5	227675	49	AE004484	Human	sapl
24	81	3.4	303819	30	AE003509	Human	sapl
25	78.8	3.4	35999	66	AC078925	Human	sapl
26	78.8	3.4	163642	52	AC073862	Human	sapl
27	78.8	3.4	252887	52	AC073527	Human	sapl
28	78.4	3.3	1326	32	PMU09766	Human	sapl
29	78.4	3.3	1326	32	SG9014	Human	sapl
30	78	3.3	4139	39	HUMPANWU	Human	sapl
31	76.8	3.3	218735	92	AC011899	Human	sapl
32	76.2	3.2	3941	12	RNLRA	Human	sapl
33	76	3.2	7218	5	I66494	Sequence	
34	75.8	3.2	1072	73	HS2IEA	Human	sapl
35	75.8	3.2	2423	11	AF109134	Human	sapl
36	75.8	3.2	6670	11	AF112980	Human	sapl
37	75.8	3.2	133661	72	U93872	Human	sapl
38	75.8	3.2	160241	35	HS885L7	Human	sapl
39	74.2	3.2	2423	11	AF112451	Human	sapl
40	74.2	3.2	2483	11	AF172453	Human	sapl
41	73.6	3.1	1345	31	PEACSPB	Human	sapl
42	73.6	3.1	220469	52	AC074307	Human	sapl
43	73.2	3.1	2479	12	RN87960	Human	sapl
44	72.8	3.1	219455	39	AC012179	Human	sapl
45	72.8	3.1	1302	32	PMKAC2IRC	Human	sapl
46	72.8	3.1	1302	32	PMKAC2IRC	Human	sapl

ALIGNMENTS

AF015297	Human	her	/protein_id="AA664404.1"
AF015298	Human	her	/db_xref="GI:7595303"
U13194	Human	herp	/translation-"NDPCCDSTWEDEDEDAEDAEDECCEGEAAAGADADAGDEDE
X83413	Human	herp	SEEPKARPSFSRMTGSRNMRATRDICRTAHNPDPDVEVDCGDPDPNLSFNEIKR
AC020678	Homo sapi		FLPNGCFIEDIILQMTNDYNDLLEDHNSYIOMLEPREGVMMHAKPITLREVEYFKSS
AR030811	Sequence		QEOIERLHVAYVELMGFYGIRLEDGEGTGVGAQYQKRFQNLMMRSHNNRLIYRIK
AR030812	Sequence		SCPELSHFHOAVLVEFLEFLETLREELRGVGLSDIFRPAVCRHROLVIAFNAE
AC023920	Homo sapi		HPRPKCFVWQDPRDKRFPKPSLHPLEGSKAKVKEESPDPGHEHSTOGRSTFEPH
AC063665	Homo sapi		SKGGGVDESDPQDVSYPQDAPGLERSGDGAAGVGEDEPPLSPKSKRKTELNR
X74358	P. carnea	Bo	EQPTTBERGQSAEYKILNIEGALSQGSRTCTQVGGDQDGEAVOICRQPLGR
AC044845	Mus muscu		VADKVRKRKRVSDGADSAVASGAGQTLALGSPABSGHFKAGHSESVEDTEBGR
AE003509	Drosophi		GPKEGTGGSPSETGGSPAGPAGDEPAASSTETPPRAGPAGADEPSPAGEPTEGLR
AC078925	Homo sapi		ACGACGDEPAETPSETETGGSPAGCTPDEDAESSETGGSPRAGPAGDEPAESSTPP
AC073863	Homo sapi		RPAGPAGDEPAESSETGGSPAGCTPDEDAESSETGGSPRAGPAGDEPAESSTPP
AC073527	Homo sapi		BASE COUNT 503 a 754 c 843 g 323 t
U09765	Plasmodiu		ORIGIN
S69014	Plasmodiu		
U05582	Human pancr		
AC011899	Homo sapi		
X83545	R. norvegic		
U66494	Sequence	14	
M29385	Herpes slm		
AF109134	Homo sapi		
AF112980	Homo sapi		
U93872	Kaposi's SA		
AL035669	Human DA		
AF172451	Homo sapi		
AF172453	Homo sapi		
U03302	Plasmodiu		
AC074307	Mus muscu		
U87960	Rattus norv		
AC012179	Homo sapi		
U001526	Plasmodiu		

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1	AF172451	Homo sapiens opioid growth factor receptor mRNA, complete cds.	AF172451	2423 bp	19-APR-2000	human.						
			AF172451	GI:7595302		human.						
						Homo sapiens						
						Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
						Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.						
						1 (bases 1 to 2423)						
						Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.						
						Cloning, sequencing, chromosomal location, and function of cDNAs						
						encoding an opioid growth factor receptor (OGFR) in humans						
						Brain Res. 856 (1-2), 75-83 (2000)						
						2 (bases 1 to 2423)						
						Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.						
						Direct Submission						
						Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,						
						Pennsylvania State University College of Medicine, 500 University						
						Drive, Hershey, PA 17033, USA						
						Location/Qualifiers						
						1. 2423						
						/organism="Homo sapiens"						
						/db_xref="taxon:9606"						
						/chromosome="20"						
						/map="20q13.3"						
						/tissue_type="placenta"						
						34. 2067						
						/note="(Met-5)-enkephalin binding protein; OGFR;						
						alternatively spliced"						
						/codon_start=1						
						/product="Opioid growth factor receptor"						

		/protein_id="AF664404.1"	
		/db_xref="GI:7595303"	
		/translation="MDPDCDSTWEDEDEDAEAEDEDEDEDEGAARADAG SEEPKARPSFSRMTGSKNMFRATDMCKITRINTPDLVLRDNGSTPLPSF FLNCGCTEDIIQNTWDNDYDLDLEDNHSTYIQMLEPVMNAPKALTLEVE QEIQERLVRAYVELMAGFYGIRLEDRETGVGRKONYOKFQNMNSHNRLRV SPEILSLEHFOALVLEFLEETLVREELGVRSALDYFMVAIRCHORSRLV HPPKCFEVMGPQDKLREKFSPLPGLSREKVEEGSPGDPHSAFOSRST SKGGKRVDEGPQSRVSDPQAGLERLOSGDEGHEDPEPLSPENSKRKRI EQPTPEGVPDASAEVYKIALNIEGALSQGSILRTGYOEVGGDPCRAVYPCR VAKRKRNRKVRDEGASAVVAGGAKOTIALASLPVSHPPKSEKNEVEE GPEKVRTPGSPSETPGSPAPADDEAPSESTPGRPAPADDEAPSESE RAGGAEDEAPTPSETPGSPAPTRDPEAPSESTPGRPAPADDEAPSESE RPGAPGADPEAPSESTPGSPAPTRDPEAPKAGEAEIQQDEVESSAKSRF	
BASE COUNT	503 a 754 c 843 g 323 t		
ORIGIN			
Query Match	100.0%;	Score 2348;	DB 11; Length 2423;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 2348;	Conservative 0;	Mismatches 0;	Indels 0; Gaps
QY	1	tagaattcagcgccgcgtgaaatttcagccgagatcagacaccccgctcgattccacc	60
Db	1	TAGATTTCAGGCGCGCTGAAATTTTCACCGAGCTGAGACACCCCGACTGACTCCACC	60
QY	61	tgaggagagagacgagagagatcgcgagagacgcgagagacagagactcagagacgcgag	120
Db	61	TGGAGGAGAGACGAGAGAGATGCGGAGAGACCGGAGGACGAGACTCGAGAGCGCGAG	120
QY	121	gcccgcgcgcgcgcgagagagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg	180
Db	121	GCCCGCGCGCGCGAGGAGACGCGGACGCGAGGGGACGAGGACGAGGAGTGGAGGAGCCGCG	180
QY	181	gcg	240
Db	181	GCGGCGGCGCGCCACTCGTTCCACTCCAGATGACAGGCTCCAGAACTGGCAGGCCACG	240
QY	241	aggagacatgctgtaggtatcgcgcacaactatccggatctggttgagagagactgcaattggg	300
Db	241	AGGGACATGTTAGGTATTCGGCACAACTATCCGGATGTGGTGGACAGACTGCAATGGG	300
QY	301	gacacgcacaacactcgagttcttaccagaaatagatccgcttccctgcgccacacgcgcgtgttc	360
Db	301	GACACGCCAAACTCGAGTTTCTACAGAAATAGATTCGCGTTTCTGCCCAACGGCTGTTTC	360
QY	361	atgtagaacattcttcacgaactctgagacggaacataatgacctctttagagaacatctacc	420
Db	361	ATTGAGGCAATTTCTTCAGAACTGGACGCGACAACTATGACCTCTTGAGCAATCACTCC	420
QY	421	tacatccagctgctggtttccctcttcgagaaacaggaatgaaactggagctggccaagccctc	480
Db	421	TACATCCAGTGGCTGTTTCCCTCTCGAAGAACAGAGATGAACTGGCAATGCCAAAGCCCTC	480
QY	481	acgcctcagggagagctcgaaggtgtttaaaagctcccaagagatcccaagagacgcgtctgtccg	540
Db	481	ACGCTCAGGGAGGTCGAGGTTTAAAAAGCTCCCAAGAGATCCAGGAGCGGCTGTGTCGG	540
QY	541	gctctaagagctcatgcttgcgccttctacaggatccgcgtctgtagaagacggagcaacggagcag	600
Db	541	GCTTACGAGCTCATGCTGCGGCTTTTACGGGATTCGCGCTGAGAGCCACGACGCGGGACG	600
QY	601	gtgggcgcgaagacaagaactaccagaagcgttccagaagaactggaactgagcgcagccaaac	660
Db	601	GTTGGGCGGAGACGAACTACCAAGACGCTTCCAGAACTGAATCTGGCCAGCCACAC	660
QY	661	aacctccgcacataacgcgcatcctctaagctgcgcgtgtgtagcttgagcctcgcgacattccag	720
Db	661	AACCTCCGCATCACACGCGATCCTCAATCGCGGCTGAGCTGAGCCCTCGAGCACTTCAG	720
QY	721	gcgcgcactgctgcgccttcttcgtgagagaaagcgtgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcg	780
Db	721	GCGGCACAGTGCCTTCTTCGTGAGAGAAAGCTGTGTGCGGCGGAGCTCCGGGGGTG	780

[illegible]

QY	1861	gaaaggagaaacgaacgcgaagaagccatctggagaccacagcccccgcgcgagaagact	1920
Db	1861	GCAAGGGACGACCCACCCAGAGACCCATGCGAGACCCAGGCCCCGCCCGCAGACT	1920
QY	1921	gcaaggagacgaacgaacgcgagaagcccatctggaaccaccaagccccagcccgagagact	1980
Db	1921	GCAAGGGACGACCAACACCCAGAGCCCATGCGAGACCCAGGCCCCAGCGCGAGACT	1980
QY	1981	acaaaggatgaacgaacccaagcgcgggagagcagcagaagtctgcaagacgcagagtgag	2040
Db	1981	ACAAGGATGAGCCACCCAAAGGGGGGAGCGACAGAGATTGCAAGAGCGCAGAGTGAG	2040
QY	2041	tctctgcgaagctctctggagaagccttaagaaagagatgcgcgttggtcttgctctcc	2100
Db	2041	TCTTCTGCCAGACTCTGGGAAAGCCTTAAAGGAAAGAGATGCCGCTGCGCTCTTGGTCTCC	2100
QY	2101	tgtccctgcgcgaaggagctggggagccttcggaactctctcgaaatcccttaaggtctgact	2160
Db	2101	TGTCCCTGCTGCAAGGGGCTGGGGGCTTCGCGAACTCTCTGGGACTCTCCCTCAAGGCTTGCT	2160
QY	2161	tctgtgacccgtgagcccatgaacccaagatgtgtgctctctctgtgggcaactaagaagcca	2220
Db	2161	TCTGTGACCCCTGACCCATGACCCACCAAGTCTGTGGCTCCTGTGGGGCCACTTATGACAGCA	2220
QY	2221	ccgaagaacgcgcgaagccttcagggaaagccaagagcctgcagaagcctctgagcgtgctg	2280
Db	2221	CGAAGAACGCCGAGGGCCCTCAGGAGAAAGCCCAAGGGCTGTGAGAAAGCCTCTGTGGCTGTG	2280
QY	2281	tgtcttcccaaccagactctccctgagccctgtctcttgttaaatgaacctctbgaagt	2340
Db	2281	TGTCTTCCCAACCAACTCTCCCTGCGGCCCTGTGTGTGTAATTAAGACCTTGTGAGGT	2340
QY	2341	gggggggagc 2348	
Db	2341	GGGGGGGG 2348	

RESULT	2
LOCUS	AF172453
DEFINITION	AF172453 2483 bp mRNA PRI 19-APR-2000
ACCESSION	Homo sapiens opiod growth factor receptor mRNA, complete cds.
VERSION	AF172453
KEYWORDS	AF172453.1 GI:7595306
SOURCE	.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE	1 (bases 1 to 2483)
JOURNAL	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
MEDLINE	Cloning, sequencing, chromosomal location, and function of cDNAs encoding an opiod growth factor receptor (OGFR) in humans
REFERENCE	Brain Res. 856 (1-2), 75-83 (2000)
AUTHORS	2 (bases 1 to 2483)
TITLE	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
JOURNAL	Direct Submission
MEDLINE	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University
REFERENCE	Dive, Hershey, PA 17033, USA
AUTHORS	Location/Qualifiers
TITLE	1. 2483
JOURNAL	/organism="Homo sapiens"
MEDLINE	/db_xref="taxon:9606"
REFERENCE	/chromosome="20"
AUTHORS	/map="20q13.3"
TITLE	/tissue-type="placenta"
JOURNAL	34.. 2127
MEDLINE	/note="[Met-5]-enkephalin binding protein; OGFR; alternatively spliced"
REFERENCE	/codon_start=1
AUTHORS	/product="opiod growth factor receptor"
TITLE	/protein_id="AA64406.1"

BASE COUNT	518	a	778	c	861	g	326	t
ORIGIN								
Query Match	96.4%	Score	2263.6	DB	11	Length	2483	
Best Local Similarity	97.1%	Pred. No.	0					
Matches 2339	Conservative	0	Matches	9	Indels	60	Gaps	1

[illegible]

QY	781	cgscagaaatgcgcctggagacttcaatgatgtccggtgtgcgtgcgcgaacaaagccgccaag	840
Db	781	CGCGCAaATGcCCTGgATaCTTATgTTGCGGTGCGGTGCAGACCAAGCGCGCAAG	840
QY	841	ctgtgtacattcgtcctgtgagacattccggcccccgctgcaagtctgtctgtgggccccaa	900
Db	841	CTGTGTACaCTTTCGcCTTGgAGaCAATTTCGCGCCCGCTGCAGaGTGCTGTGGGGGCCCAa	900
QY	901	gaacaagtctcgagaggttccaagcccagctctctctgcgcatccgctctcgaggtctcaagag	960
Db	901	GAaCAAGCTGCGGAGaGTTCAAGCCCAAGCTCTCTCTCCCACTCCCTGTAGAGGCTTCAGAGAG	960
QY	961	gtgtagagagagaaagaaacccccggggagacccccgcacagagagagcaacaaagtctagacc	1020
Db	961	GTGTAgaGAGAAAGAAAGaACCCCGGGGAGaCCCGACCAAGAGCCaACCACTTAGGTGTGAAC	1020
QY	1021	tgttggccaagacataagcaagaggt	1080
Db	1021	TGTGTAGCaGAGCAATTAAGaAGGtGGGGGAGAGGTGTGAGAGAGGGGCCCAAGCAAGAGAC	1080
QY	1081	gttgtagccccctgagtgctgt	1140
Db	1081	GTGTAgaGCCCCAGATGTGAGAGaCCCTGTGAAGAGAGCCAGGGGAGATGAGCAAGGGGGCAAC	1140
QY	1141	ggggagaagataagccggggagcccttaagcccccaagagaaagcaagagagctgtgactgt	1200
Db	1141	GGGGAAGATaAGCGCGAGaCCCTTAAGCCCAAGAGaAGCAAAAGAGAGaAGCTGTGAAGCTG	1200
QY	1201	agccggtcggtgagcagccgcccccaagacagccctcagaagtgtcctcagaagtgtgaaag	1260
Db	1201	AGCCCGCGGAGaCGaCCCGCCACAGGGCCAGGGCCCTCAAGATGTCTTAGAGGTGTGAAGAG	1260
QY	1261	atcgctctgaatttggtaggggtgtgtccctcaagccaaaggagcagctcaagacggtagaacag	1320
Db	1261	ATCGCTCTGAATTtTGAGAGGGGTGTCCCTCAAGCCAGGCaAGCTTCAAGAACGGGAGaCCAG	1320
QY	1321	gaagtgtggctgtcaggaaccccttggtagagcagtgtcagcccttgcgcgcaaaccccttggagcc	1380
Db	1321	GAAGTGGCGGTGTCAAGaCCTTGGGAGAGCaGTGAACCTCTGCGCCCAaACCTGGGAGaCC	1380
QY	1381	aggtgtgtgcgcgcacaaagtgtgaagaaagcggtagaagtgtgatagagtgctctggtagacagt	1440
Db	1381	AGGGTGTGCGCGaCAAGGTGTGAAGAAAGCGaAGAGTGTGAAGAGGTACTGTGGGAGaCACTCT	1440
QY	1441	ggcgtgtgccaagtgtgtgtgcaccagaccttggcccttgcgcggtgtccctctggccccatcgag	1500
Db	1441	GGCGTGTGCGaGTGTGTGTGTGCaGACCTTGGCCCTTGGCGGGTCTCCCTGTGCCCACTCGGGG	1500
QY	1501	caccccaagagctgtgacacagttgaaacgggtgttagagagacacagaaagtctgaaacgggg	1560
Db	1501	CACCCCAAGAGCTGTGAACaCACTGTGAAGaCGGGGTGTGAAGAGaCAGCAAGAGGTCTGAACGGGG	1560
QY	1561	cccaaaagaagttaaccccttggtagagcccatctggtagacccccagggccccagccagagagact	1620
Db	1561	CCCAaAGaAGSTaACCTTGGGaAGCCCATGTGAGAGaCCCAAGGCCCAAGCCCAAGCaGACT	1620
QY	1621	gcagagggagac-----	1629
Db	1621	GCAGAGGCGAGaGCCaAGCaAGaCCCATGTGAGAGaCCCAAGGCCCAAGCCCGGCaGAGACT	1680
QY	1680	-----gagccaggtccgagagagcccatctcgtagagaccccaggtcccccgcacagagact	1680
Db	1681	ACAAGGATGAGCCaAGCCAGCaGACCCCATGTGAGAGaCCCAAGGCCCCCGCCCGCaGAGACT	1740
QY	1681	gcagagggacagccagcccgtagagcccatctggtagacccccagaggtctctgcgcggtagagact	1740
Db	1741	GCAAGGAGaCAAGCCaAGCCaAGaCCCATGTGAGAGaCCCAAGGCCCCCGCCCGCaGAGACT	1800
QY	1741	gcagagggacagccagcccgtagaccccatctgtagacccccagaggtcccccagcccgtagagact	1800
Db	1801	GCAAGGAGaCAAGCCaAGaATTCaCTGTGAGAGaCCCAAGGCCCCCGCCCGCaGAGACT	1860

QY	1801	aaagagatgaagccagccagagcccatctggagaccacccagagcccgcccgccgagagact	1860
Db	1861	ACAAGGATGAGCCAGCCAGCCAGCCCATGAGAGACCCAGGCCCCCGCCGGCAGAGACT	1920
QY	1861	gcagggagacagccagccagagcccatctggagaccacccagagcccgcccgccagagact	1920
Db	1921	GCAGGGAGACGACCCAGCCAGGCCCATGAGAGACCCAGGCCCCCGCCGGCAGAGACT	1980
QY	1921	gcagggagacagccagccagagcccatctggagaccacccagagcccgcccgccagagact	1980
Db	1981	GCAGGGAGACGACCCAGCCAGGCCCATGAGAGACCCAGGCCCCCGCCGGCAGAGACT	2040
QY	1981	aaagagatgaagccagccagagcccgaggagggagagagagttgcagagagccagagttgag	2040
Db	2041	ACAAGGATGAGCCAGCCAGCCAGCCGAGGAGGAGCAGAGAGATTGACAGACGACAGAGTTGAG	2100
QY	2041	tctctgcgaagctctggagagcccttaagaaagagatggccgtctgtgtgtgtgtgtgtgt	2100
Db	2101	TCTTCTGCCAATCTTGGGAAAGCCTTAAGGAAAGAGTGCCTTCGGCGCTTGTGTGTGTGT	2160
QY	2101	tgtccctgctgcaagggctgtgggacctccgagctgtctgtcgagactccctcaagcttgt	2160
Db	2161	TGTCCCTGCTGTCAGAGGGGCTGGGGGCTCCGAGAGCTGTGTGGGGCTCCCTCAAGCTGTGT	2220
QY	2161	tgtgtaccctgttaacccatgagaccacaagtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2220
Db	2221	TGTGTACCCGTATACCATATACCAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2280
QY	2221	ccagagaccgagagccctccagagagagaccagagctgtgagagagctctctgtgctgtgt	2280
Db	2281	CCAGAGACCCGACGAGGCTCTCAGAGGAGCCCAAGGCTGTGACAGAGCTCTGTGCTGTGGCT	2340
QY	2281	tgtctcccccacccagctctccctctgcgcctctgtctgttaattgacccctctcgaat	2340
Db	2341	TGTCTTCCCAACCAACGCTCTCCCTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2400
QY	2341	ggggggggcgcg 2348	
Db	2401	GGGGGGGCGC 2408	
RESULT	3		
LOCUS	AF109134	2423 bp	10-JAN-1999
DEFINITION	Homo sapiens 7-60 mRNA, complete cds.		
ACCESSION	AF109134		
VERSION	AF109134.1	GI:4139227	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 2423)		
AUTHORS	Liu,J., Brewton,R.G., Takanosu,M., Wood,B.M. and Mayne,R.		
TITLE	Cloning of 7-60: A human gene from chromosome 20q13.3 which encodes a novel intracellular protein motif repeated seven times		
JOURNAL	2 (bases 1 to 2423)		
REFERENCE	Direct Submission		
AUTHORS	Submitted (23-NOV-1998) Cell Biology, University of Alabama at Birmingham, 1670 University Blvd. VHS02, Room 605, Birmingham, AL 35294, USA		
JOURNAL			
FEATURES	Location/Qualifiers		
SOURCE	1..2423		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="20"		
	/map="20q13.3"		
	/note="approximately 3kb upstream of COL9A3"		
	206..2062		
	/codon_start=1		
	/product="7-60"		
CDS			

Query Match	Best Local Similarity	96.1%	Score 2256.6	DB 11	Length 2423
Matches 2201	Conservative	99.1%	Pred. No. 0	Mismatches 19	Indels 3
				Gaps 3	
27	gccagcatalygcagaccccgactgcgactccactctggaggagagacagagagatgcgga				
23	gccgagcattgcagaccccgactgcgactccactctggaggagagacagagagatgcgga				
QY	ggagcgcgagagacgagagactgcgagagacgagagccgcgcgagagagacgagagcgc				
Db	ggagcgcgagagacgagagactgcgagagacgagagccgcgcgagagagacgagagcgc				
QY	aggagagcagagacgagagagtcgagagacgagagccgcgcgagagagcgcgcgcacgtc				
Db	aggagagcagagacgagagagtcgagagacgagagccgcgcgagagagcgcgcgcacgtc				
QY	cagaaatgcagaggtccagaaacttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	cagaaatgcagaggtccagaaacttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	cgatccgagatctcggtggagagagactgaaatggagacgcgcgcgcgcgcgcgcgcgcgc				
Db	cgatccgagatctcggtggagagagactgaaatggagacgcgcgcgcgcgcgcgcgcgcgc				
QY	aaatgagatccgcttccttcgc				
Db	aaatgagatccgcttccttcgc				
QY	ggaacacatgatgcctctctttagagagacatcctctacatcagttgctgttcccttcgcg				
Db	ggaacacatgatgcctctctttagagagacatcctctacatcagttgctgttcccttcgcg				
QY	agaacacagagatggaactggcatggcgaagccctctacgctcagggaggtcgaggttttaa				
Db	agaacacagagatggaactggcatggcgaagccctctacgctcagggaggtcgaggttttaa				
QY	aagctccagagagatccagagagcgcgttcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	aagctccagagagatccagagagcgcgttcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	aagctccagagagatccagagagcgcgttcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	aagctccagagagatccagagagcgcgttcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
QY	agtcgcggttgtagcttgagcctcgcgcacatccagagcgcgcgcgcgcgcgcgcgcgcgcgc				
Db	agtcgcggttgtagcttgagcctcgcgcacat				

[illegible]

Db	1881	CATGGGAGACCCACAGGCCGCCGCCGCGCAGGACTTCAGAGGGGACGAGCCAGCAGAGCC	1940
OY	1946	catgagagaccaccaagcccccagcccgagagacactcaagagatagcccaacgaagcgg	2005
Db	1941	CATGGGAGACCCCAAGGCCGCCGCCGCGAGACTTACACAGGATATAGCAGCCAAAGGGG	2000
OY	2006	gagagagacagagatttcagaaagcagagatgtagtctcttcagcaagtcgtgaaagcct	2065
Db	2001	GGAAGGAGCAGAGTTTCCAGGAGCGAGAGTGGAGTCTTCTGSCAAGTCTGGAAAGCTT	2060
OY	2066	aaggaagagagttgcgcgctcggtcttgctctctctctcctgctcgtcaggggctggggcc	2125
Db	2061	AAGGAAGAGAGTGCCTGCGCTCTTGGCTCTCTGTCCTCCGTCGACGAGGGGCTGGGGCC	2120
OY	2126	tcggagactgctgagactatccctcaagtgctgcttcgttcgtgacccgtgacccaatgacccac	2185
Db	2121	TCCGGAGCTCTCTGGGGCTTCCTCCACAGGCTTCGCTTCGTGACCCGTGACCCGATACCCAC	2180
OY	2186	agtctgtgacctcctgttgaggccactatgacgacacagaagccgagagccctcagaagga	2245
Db	2181	AGTCTGTGGCTCTCTGTGGGGCACTTATACACACCACGAAAGCCGCGAGGCCCTTCAAGGA	2240
OY	2246	agcccaagagcctgcagaagacctctcgtgcctgtgtcttcaccacaacagcttccct	2305
Db	2241	AGCCCAAGGCGCTCAGAAAGCCTCTGGCTGGCTGTGCTTCCACACAGACTCTCCCT	2300
OY	2306	gagccctgtctctgttaatatgacctctctgagctggggagcg	2348
Db	2301	GGCGCCCTGCTTTGTAAATGACCCCTTCTGGAGGGGGGGCG	2343

RESULT	4
LOCUS	AF172452
DEFINITION	AF172452 Homo sapiens oploid growth factor receptor mRNA, complete cds.
ACCESSION	AF172452
VERSION	AF172452.1
KEYWORDS	GI:7595304
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 2363)
TITLE	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J. Cloning, sequencing, chromosomal location, and function of CDNA encoding an oploid growth factor receptor (OGFR) in humans
JOURNAL	Brain Res. 856 (1-2), 75-83 (2000)
MEDLINE	20143540
REFERENCE	2 (bases 1 to 2363)
AUTHORS	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy, Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA
FEATURES	location/Qualifiers
SOURCE	1..2363
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.3"
	/tissue-type="placenta"
	34..2007
	/note="(Met-5)-enkephalin binding protein; OGFR; alternatively spliced"
	/codon_start=1
	/product="oploid growth factor receptor"
	/protein_id="AAFG4405.1"
	/db_xref="GI:7595305"
	/translation="MDDPDCSTWEEDEBADEDEDCDEGAAGRADAGDEDEE SEEPARPSFSRMTGSRMWRATROMCARRHYPDLVEDDCGDPNPISFYNEITR LENGCEIIVLQNTWIDNLDLEDNHSYIOWLPRLRPGVWMAKPLTLREVFKSS QEIORERIVAYELIMGEGYIRLEDGSGITVGRGAQNYOKRFONLWNRSHNNRIIRLIK
CDS	

SPCELSLEHFOAPLVREFLEETLVRELPGVRSQALDYFMEAVCRHOROLVHFAWE
HFRPRCKRVWEPQDKLRFKRPSPLEPHLEGSRKVEEESPDPDHEATOGRTGPEH
SKGGRVNDEGPSPRSVERPODAGPSPLEPSRSGDREPSRSPLEPSRKRKRLISRR
EQPTEBPQASAEYERKALNIEGALSQSLRSTQDVGODPEANQPRQPIEGAR
VADVVRKRVKDEGADSAVAGGQTLALAGSAPGSHKAGSENGVEDTEGRT
GPKRGTPSPSPETBPSPAGADBPAPSPETBPSPAGADBPAPSPETBPSP
AGPRDEPAESPSPETBPSPAGADBPAPSPETBPSPAGADBPAPSPETBPSP
SPAGTRDEPAKGAELQDAVESSAKSKSP"

BASE COUNT 488 a 729 c 825 g 320 t 1 others
ORIGIN

Query Match 94.2%; Score 2212.8; DB 11; Length 2363;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2284; Conservative 1; Mismatches 3; Indels 60; Gaps 1;

QY 1 tagaatcagcgccgctgaattctacgcagcatgagaccccgactgcattccacc 60
DB 1 TAGAATTGACGGCGCGCTGAATTTAGCCGACATGAGACCCGACTGCATCTCAC 60

QY 61 tgggagagagagagagagatgcgagagacgagagagagagagagagagagag 120
DB 61 TGGGAGAGAGAGAGAGAGATGCGAGACGCGAGAGAGAGAGAGAGAGAGAGAG 120

QY 121 gccgc 180
DB 121 GCCG 180

QY 181 gc 240
DB 181 GCGGCG 240

QY 241 agggagagatgtagatctcgcacacatctccgagatctgtgtgaacagagac 300
DB 241 AGGGAGAGATGTAGATCTCGCACACATATCCGATCTGGTGAACAGACATG 300

QY 301 gacacgcac 360
DB 301 GACACG 360

QY 361 attgagagacatctctcagaacacacacacacacacacacacacacacac 420
DB 361 ATTGAGGACATCTCTCAGAACACTGAGACACTATGACCTCTTGAGACATC 420

QY 421 taccatccagtgctgttctctctgagagacacagagatgaactgtgcacgccc 480
DB 421 TACATCCAGTGCTGTCTCTCTGAGAACAGGATGCTGATGCCAAAGCCCTC 480

QY 481 agcgtcagagagagagagagtggttttaaaagctccagagagatccagagag 540
DB 481 ACGCTCAGGAGAGGTGAGGTGTTTAAAGCTCCAGAGATCCAGAGCGGCTGT 540

QY 541 gctcagagacatgctgtggtctctcagagacacgctgtgagagacacagag 600
DB 541 GCCTAGGACACTGCTGTGCTTCTAAGGATCCGCTGAGAGACCGAGACG 600

QY 601 gtggagcagagacagaaactacagaaagcgctccagaaactgaactgtgcag 660
DB 601 GTGGGCGCGGACAGAACTACAGAAAGCGCTTCCAGAACTGAACGTGGCG 660

QY 661 aaccccgagatagacagacatctcagagtgccgtgtgagctgaagctgagac 720
DB 661 AACCTCCGATACAGGACATCTCAAGTGCCGTGTGACCTGACCTCAGACTTC 720

QY 721 gccgcaactgtgctctctctcagagagagacgctgtgagcgcgcgcgcgcgc 780
DB 721 GCGGCACTGTCTCGCTTCTCTGAGAGAGAGCTGTGCGGCGGAGAGTGC 780

QY 781 cggcagagatgcccgtggaactactcatgttcgcggtgagctgcgcagacag 840
DB 781 CCGGAGAGAGGCGCTGTGACTACTTCAATGTTCGCGTGCAGACCCAGGCG 840

QY 841 cgggtgagacatctgcctggagacatccgcgcgcgcgcgcgcgcgcgcgc 900

DB 841 CTGGTGCACTTCGCTGGAGAGACTTCGCGGCCGCTCAGAGTTGTGTGGGGCCCA 900

QY 901 gacaaagctgcggaggttcaagaccagctctctgcgcacatccgctcagaggtc 960
DB 901 GACAAAGCTGCGGAGGTTCAAGCCCGCTCTCTGCGSCATCCGCTCGAGGGCTCC 960

QY 961 gtggag 1020
DB 961 GTGGAG 1020

QY 1021 tctggcagagacatagcaagaggtgtgggagcaggtgtgagcagagggcccc 1080
DB 1021 TCTGGCCAGAGACATAGCAAGGTGTGGGAGCAGGTGTGAGAGAGGGCCCCAG 1080

QY 1081 gtgagaccccaagatctgcggagcccttgagagagacagggagatgagcgag 1140
DB 1081 GTGAGACCCCAAGATCTGCGGAGCCCTTGAGAGAGACCCAGGGGATGAGCG 1140

QY 1141 gggagagatagcgagcccttaagccccaagagagcaagaaagagagagagctg 1200
DB 1141 GGGAGAGATAGCGAGCCCTTAAGCCCAAGAGACCAAGAGAGAGAGAGAGAG 1200

QY 1201 agc 1260
DB 1201 AGCG 1260

QY 1261 atcgcctgaatttgagaggggtgtgcctcagcagagacacacacacacacac 1320
DB 1261 ATCGCTGGAATTTGAGAGGGGTGTGCTCTCAGCCAGGCGACCTCAGAGAG 1320

QY 1321 ggaagtggcggtcagagacccctggagagagcagtgcagccctgcgcgaaccc 1380
DB 1321 GGAAGTGGCGGTCAAGACCCCTGGAGAGAGCAGTGCAGCCCTGCGCAACCC 1380

QY 1381 aagggtgcgcgaaggtgagagagcagagagagagagagagagagagagag 1440
DB 1381 AAGGTGCGCGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

QY 1441 ggcgtgagcagtggtgtgcagacacacacacacacacacacacacacacac 1500
DB 1441 GCGGTGCGCAGAGGTGTGCGCACACTTGCGCTTGGCGGCTCCCTGCGCA 1500

QY 1501 caaccacagagcgtgcagacaggtgagagacgggtgtgagagagagagagag 1560
DB 1501 CACCCAGAGCCTGAGACAGAGTGAAGACGGGTTGAGAGAGACAGAGAGTGA 1560

QY 1561 cccaaagagagtaacccctggagacacacacacacacacacacacacacac 1620
DB 1561 CCCAAGAGAGTAAACCCCTGGAGACCATCGAGACCCAGGCCCAAGCCAG 1620

QY 1621 gcaagggagcagacagccagagagacacacacacacacacacacacacac 1680
DB 1621 GCAAGGGAGACAGACCCAGAGAGCCCATCGAGACCCAGGCC----- 1680

QY 1681 gcaagggagcagacagccagagagacacacacacacacacacacacacac 1740
DB 1681 -----CCGCCGCGAGAGACT 1740

QY 1741 gcaagggagcagacagccagagacacacacacacacacacacacacacacac 1800
DB 1741 GCAAGGGAGAGACCCAGAGAGCCCATCGAGACCCAGGCCCGCGAGAGACT 1800

QY 1801 acaagggagtgagcagcagacacacacacacacacacacacacacacacacac 1860
DB 1801 ACAAGGGATGAGACCGAGAGAGCCCATCGAGACCCAGGCCCGCGAGAGACT 1860

QY 1861 gcaagggagcagacagccagagagacacacacacacacacacacacacacac 1920
DB 1861 GCAAGGGAGAGACCCAGAGAGCCCATCGAGACCCAGGCCCGCGAGAGACT 1920

QY 1921 gcaagggagcagacagccagagagacacacacacacacacacacacacacac 1980

Db 1861 GCAGGGAGACGACGCCAGCCATCGAGAACCCCAAGCCCGAGAGACCT 1920
QY 1981 acaaggatgatgcagcccaaggcggggagggcagagatgtcaggagcagagagtgagg 2040
Db 1921 ACAAGGATGATGAGCCAGCAGAGCGGGGAGAGCAGAGAGTTGACAGACGAGAGTGAG 1980
QY 2041 tcttcgcacagcttggaagccttaaggaagaagtgcccgctggcgttctgtctcc 2100
Db 1981 TCTTCTGCAAGTGTGGAAAGCCTTAAGGAAAGAGTCCCGCTGCGCTTGTGCTCC 2040
QY 2101 tcttcctgtcaggggagctgcgcgcagagctgtcgcgagctccctcagagctctgc 2160
Db 2041 TGTCCCTGCTGCAAGGGGCTGGGGCTCCGAGCTGCTGCGGGCTCCCTCAGGCTCTGCT 2100
QY 2161 tctgtagccgtgacccatgacccacagctgtgcctctgttgaggccactatagcagca 2220
Db 2101 TCTGTGACCGGTGACCATGACCCACAGTGTGCTGCTGCTGCGGCTCAGTATGACAGCA 2160
QY 2221 ccagaagccgagggcgcctcagaaggagcctcagagcctcctcgtgctgctg 2280
Db 2161 CCGAAGCCGCGAGGCCCTCAGGGAAGCCCAAGGCTGCAAGAGCCTCCTGCTGCTG 2220
QY 2281 tgcctccacccagcctcctcctgcgcctgtctgttgaatgacccctctgagct 2340
Db 2221 TGTCTTCCACACCCAGCTCTCCCTGCGCCCTGCTTGTGTAATGACCTTCTGAGT 2280
QY 2341 gggggagcg 2348
Db 2281 GGGGGGCG 2288

RESULT 5
HS885L7
LOCUS HS885L7 160241 bp DNA PRI 09-JUL-2000
DEFINITION Human DNA sequence from clone RP5-885L7 on chromosome
ACCESSION AL035669
VERSION AL035669.43 GI:8979786
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 160241)
Smith M.
Direct Submission
Submitted (09-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 8, 2000 this sequence version replaced gi:8919619.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone configs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP5-885L7 is from the library RPc1-5 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.bufileo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-885L7 The true left
end of clone RP4-563E14 is at 140192 in this sequence. The true
right end of clone RP11-93B14 is at 44512 in this sequence.
FEATURES
Source
Location/Qualifiers
1..160241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.2-13.33"
/clone="RP5-885L7"
/clone_lib="RPc1-5"
2..64
/note="LIM9 repeat: matches 6210..6254 of consensus"
65..353
/note="Alu10 repeat: matches 1..287 of consensus"
354..397
/note="LIM9 repeat: matches 6254..6302 of consensus"
4196..5157
/note="12 copies 80 mer 61% conserved"
complement(4223..5153)
/note="Tandem repeat. region contains two forced joins and
single clone. Assembly consistent with restriction
digest."
4229..5114
/note="443 copies 2 mer cc 59% conserved"
4231..5166
/note="234 copies 4 mer tccc 60% conserved"
4240..5151
/note="19 copies 48 mer 61% conserved"
4248..5159
/note="38 copies 24 mer 61% conserved"
4264..5163
/note="15 copies 60 mer 61% conserved"
4286..5152
/note="17 copies 51 mer 59% conserved"
4308..5171
/note="27 copies 32 mer 61% conserved"
4317..5156
/note="21 copies 40 mer 62% conserved"
4344..5142
/note="17 copies 47 mer 59% conserved"
4348..5145
/note="42 copies 19 mer 60% conserved"
4431..5144
/note="21 copies 34 mer 60% conserved"
4456..5148
/note="21 copies 33 mer 60% conserved"
4486..5151
/note="9 copies 74 mer 69% conserved"
4530..5087
/note="18 copies 31 mer 71% conserved"
4533..4822
/note="5 copies 58 mer 77% conserved"
4720..5137
/note="11 copies 38 mer 61% conserved"
4833..5064
/note="4 copies 58 mer 79% conserved"
5051..5164
/note="19 copies 6 mer tctctc 63% conserved"
5934..6864
/note="LIM4 repeat: matches 4509..5488 of consensus"
6974..7049
/note="4 copies 19 mer 80% conserved"
7224..7451
/note="6 copies 38 mer 98% conserved"
7892..8074
/note="MER68a repeat: matches 68..210 of consensus"
8674..8966
/note="Alu1 repeat: matches 5..296 of consensus"
10876..10962
/note="MIR repeat: matches 86..162 of consensus"

```
repeat_region 11974..12025
/note="13 copies 4 mer ctca 73% conserved"
repeat_region 14394..14478
/note="5 copies 17 mer 77% conserved"
repeat_region 15824..15891
/note="17 copies 4 mer tccc 67% conserved"
repeat_region 16266..16577
/note="AluSq repeat: matches 1..313 of consensus"
repeat_region 16578..16866
/note="17 copies 17 mer 65% conserved"
repeat_region 16581..16844
/note="11 copies 24 mer 68% conserved"
repeat_region 16582..16891
/note="10 copies 31 mer 66% conserved"
repeat_region 16583..16879
/note="9 copies 33 mer 66% conserved"
repeat_region 16585..16840
/note="16 copies 16 mer 66% conserved"
repeat_region 16586..16875
/note="5 copies 58 mer 72% conserved"
repeat_region 16590..16874
/note="15 copies 19 mer 64% conserved"
repeat_region 16593..16872
/note="7 copies 40 mer 67% conserved"
repeat_region 16598..16837
/note="5 copies 48 mer 73% conserved"
repeat_region 16603..16842
/note="3 copies 80 mer 77% conserved"
repeat_region 16638..16901
/note="44 copies 6 mer aaaga 62% conserved"
repeat_region 16664..16885
/note="3 copies 74 mer 78% conserved"
repeat_region 16670..16904
/note="5 copies 47 mer 77% conserved"
repeat_region 16848..16899
/note="13 copies 4 mer gaag 100% conserved"
repeat_region 17287..17411
/note="MER5A repeat: matches 35..165 of consensus"
repeat_region 17447..17486
/note="10 copies 4 mer ccat 92% conserved"
repeat_region 18303..18502
/note="5 copies 40 mer 66% conserved"
repeat_region 18307..18498
/note="4 copies 48 mer 67% conserved"
repeat_region 18317..18496
/note="3 copies 60 mer 74% conserved"
repeat_region 18319..18486
/note="7 copies 24 mer 66% conserved"
repeat_region 18320..18495
/note="44 copies 4 mer acac 64% conserved"
repeat_region 18321..18496
/note="11 copies 16 mer 64% conserved"
repeat_region 18325..18486
/note="9 copies 18 mer 66% conserved"
repeat_region 18327..18486
/note="8 copies 20 mer 66% conserved"
repeat_region 18497..18546
/note="25 copies 2 mer ac 72% conserved"
repeat_region 19329..19433
/note="LTR16C repeat: matches 286..387 of consensus"
repeat_region 19455..19590
/note="4 copies 34 mer 88% conserved"
repeat_region 19457..19604
/note="2 copies 74 mer 81% conserved"
repeat_region 19546..19625
/note="4 copies 20 mer 76% conserved"
repeat_region 19839..19954
/note="2 copies 58 mer 87% conserved"
repeat_region 20020..20103
/note="MIR repeat: matches 51..137 of consensus"
repeat_region 20704..20805
/note="3 copies 34 mer 90% conserved"
repeat_region 20710..20811
```

```
repeat_region /note="2 copies 51 mer 90% conserved"
20722..20805
repeat_region /note="5 copies 17 mer 89% conserved"
21349..21410
repeat_region /note="2 copies 31 mer 93% conserved"
21584..21654
repeat_region /note="LTR16A repeat: matches 88..158 of consensus"
21779..21852
/note="Alu repeat: matches 238..311 of consensus"
24787..24847
/note="MER91B repeat: matches 1..63 of consensus"
24855..25064
/note="MLT1D repeat: matches 107..348 of consensus"
25853..26065

Query Match 71.9%; Score 1688.8; DB 35; Length 160241;
Best Local Similarity 99.0%; Pred. No. 7.1e-241;
Matches 1699; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 633 ccagaacctgaactggcgagccacacaaactccgcatcacacagcatctcaagtgc 692
|||
Db 61392 CCGGATCTCTCGACAGGGCGACCCACAACCTCCGATCACAGCATCTCAAGTGC 61451

QY 693 gttgagctgagcctcgagcattccaggcgccactgttcgcttctcttgagagac 752
|||
Db 61452 GGTGAGCTGGGCTCCAGACCTTCCAGGCGCCCTGTGCTTCTTCTTGAGAGAC 61511

QY 753 gctgtgctggcgagagcttgcggggtgtgcagagtgctctgactacttcatgttcg 812
|||
Db 61512 GCTGTGGGGGGGAGCTGCCGGGGGTGCGGCAAGTGCCTGTGACTTCACTTATTCGCG 61571

QY 813 cgtgcgtgcgagacacagcgccgcagctgtgtacctgcctgtgagcaacttcggcc 872
|||
Db 61572 CGTCCGTGCGGACACCAAGCGCCCGCAGCTGTGCTGCTGCGGAGACATTCGCCGC 61631

QY 873 ccggtgcaagtgtgtgtggggggcccaagaacgtcggaggttcaagcccaactctc 932
|||
Db 61632 CCGCTCAAGTTCGTCTGTGGGGGCCCCAGACACACTCTCGAGGTTCACGCCACTTCT 61691

QY 933 gccgcatcgcgtcagaggtctccagaaagtgtgagaggaagcccgagaccgcga 992
|||
Db 61692 GCCCATCCGCTCGAGGGCTCCAGGAAGTGTGAGAGGAAGAACCCCGGGGACCCCGA 61751

QY 993 ccagagggccagaccccaagggtcgacctgtggccagacatagcaaggttggggcag 1052
|||
Db 61752 CCACGAGGCGCCAGCACCCAGGGTCCGACTGTGGCCAGACATAGCAAGGTGGGGCAG 61811

QY 1053 gttgagcagagggggcccaagccagcagcgtgtgagcccccagatccggaccctggagag 1112
|||
Db 61812 GGTGAGCAGGGGGCCCAAGCCAGAGGTGTGAGCCCAAGAGATCGGAGCCCTTGTGAGG 61871

QY 1113 gagccagggggtatgagcagaggggcccacggggaagatagcccagagcccttaagccca 1172
|||
Db 61872 GAGCGAGGGGATATGAGCAGAGGGGCCACGGGGAAGATAGCGCGGACCTTAAGCCCAA 61931

QY 1173 agagaagcaagaagaagcgtgagcttgagccggcgaggagcagccgacacagacagcagg 1232
|||
Db 61932 AGAAGAGCAAGAGAGAGAGAGCTGAGCTGAGCCGCGGAGAGCGGCCACAGAGCCAGG 61991

QY 1233 cccccaagatccctcaagagtgaggaagatgcctctaatttgaagggtgtgcccacag 1292
|||
Db 61992 CCTTCAGATGCTCTCAAGAGTGTGAGAGATGCTCTGAATTTGAGAGGGGTGCTCCACAG 62051

QY 1293 ccaggcgagcctcagagcaggggagcccaagagtggtggtcagagaccccttgggagagcagt 1352
|||
Db 62052 CCAAGGCGAGCTTCAGAGAGGGGAGCCAGAAAGTGGGGGTGAGACCTCGGAGGAGT 62111

QY 1353 gcaagccttgcggccaaccccttggagccaggtgtgcccgaaggttggagagcggaagaa 1412
|||
Db 62112 GCAAGCCTGCGCCCAACCCCTGAGGAGCGAGGTGCGGACAAAGGTGAGAGCGAGAGAA 62171

QY 1413 gttgatgaaggtgtctggagacagtgtcggtggtggcaggtgtgttccagacacttggc 1472
|||
```


[illegible]

RESULT	6				
LOCUS	AF112980	AF112980	6670 bp	DNA	PRI
DEFINITION		Homo sapiens	7-60 gene,	complete cds.	10-JAN-1999
ACCESSION		AF112980			
VERSION		AF112980.1	GI:4139271		
KEYWORDS		.			
SOURCE		human.			
ORGANISM		Homo sapiens			
		Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 6670)
TITLE	Takanosu, M., Liu, J. and Mayne, R.
JOURNAL	Genomic Structure of Human Gene 7-60
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 6670)
TITLE	Takanosu, M., Liu, J. and Mayne, R.
JOURNAL	Direct Submission
REFERENCE	Submitted (10-Dec-1998) Cell Biology, University of Alabama at Birmingham, VH 302, Room 605, Birmingham, Alabama 35294, USA
FEATURES	Location/Qualifiers
SOURCE	1..6670
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.3"
	join(<206..268,906..984,2235..2313,3147..3213,4125..4274
	4891..>6309)
	/product="7-60"
	join(206..268,906..984,2235..2313,3147..3213,4125..4274,
	4891..6309)
	/codon_start=1
	/product="7-60"
	/protein_id="A40374.1"
	/db_xref="GI:4139272"
	/translation="MTGSRNMRATDMCRVHNYPDLVERDCNGDTPNLSEYNEIRF
	LPNCFEDILQWNTDYLLEDNHSYIQLFPLRPGVWMAKPLTIREVEFKSSO
	EIOELRYLVAHNAHGLRDPAGPRHGGSPSELDEALPEELRSNNIRILRLKS
	LGELGLEHPOAPLVRFLPESLYVRRELPGVRSALDYTFMFAVCCROROLVFAMWH
	FRPRCKFWGPDQLRKRFPSSLPHPLEGRKVEEESPDPDHEASTGRTCPGHS
	KGGVRDEGPQPSRVSVEPDQAPLERQSGDAGHGDRPEPLPKSKRKLRSRE
	QPPREPQPGASVEKIALNIEGICALSGSLRTGTQEVGGODPGEAVOCROPLGARV
	ADAKRKRYDEGASAAVASGQOTLALAGSPASGHPKAGHSNEDTEGRGTG
	PKGTGSPSETPGSPAGPADEPAESPSETPGPRPAGAGDEPAESPSETGPRPA
	GPAGDEPAKPSPTPGSPAGPRTRDPAESPSETPPRRAPRAGDDEPAESPSETPPPR
	PAGPAGDEPAESPSETPGSPAGPRTRDPAESPSETPPRRAPRAGDDEPAESPSETPPPR
BASE COUNT	1367 a 1998 c 2169 g 1136 t
ORIGIN	
Query Match	71.6%; Score 1680.8; DB 11; Length 6670;
Best Local Similarity	98.7%; Pred. No. 4.2e-239;
Matches 1694; Conservative	0; Mismatches 22; Indels 0; Gaps 0;
QY	633 ccaggaacctgagctcggcgacgaccccaacaacctccgcatcaacgacgactctcaagtggcc 692
DB	4875 CCCGATCTCTCGAGGCGACCCACACACACTCCGATACACGACTCTCTCAAGTGGCT 4934
QY	693 gctgtgagctgagcctcgagcaacttcagagcgagcaactgtgctcttctcttggagagac 752
DB	4935 GGGTGAGCTGGCCTCGAGCACTTCCAGGGGCGCCTGGTCGGCTTCTCGAGGAGAG 4994
QY	753 gctgtgtgggggagagcttgcgggggtgtcgagcagagagcccttgactactatgttcgc 812
DB	4995 CTTGCTGGGCGGAGAGCTGCGGGGGGTGCGGCAAGTGCCTTGACTACTTCAAGTTTCGC 5054
QY	813 cgttcgctgcggaacacgacgcccagctgtgtcaacttcgcttcgggaggaacttcggcc 872
DB	5055 CGTGGGCTCGGAGAACACAGGCGCCGACGCTGTGTCACTTGGCTGGAGGACTTCCGGCC 5114
QY	873 ccgctgcgaagtctgtctgtggggccccaagaagaactctgcggaagtlcaagcccactctc 932
DB	5115 CCGCTGCAAGTTCTGTGTGGGGGCCCCCAAGCAAGCTCGGAGGTTCAAGCCAGCTCTCT 5174
QY	933 ggcgcatacgttcgtagaggtcccaaggaaagtctgggggaggaagaccgccgggacccgga 992
DB	5175 GCGCCATCTCGTCCAGGGGCTTCCAGGAAGTGTGAGGAGGAAGAACGCCCGGAGCCCGA 5234
QY	993 ccaagagccagacacccagagctgtgagccttgggcagagcctagcaaggttggggcag 1052
DB	5235 CCAAGAGCCAGACCCAGAGGCTCGGACTTGGGCCAGAGCATAGCAAGGCTGGGGCAG 5294
QY	1053 ggtgagcagaggggcccagacagagacgttggagcccaagatgtcgggagaccccttgagag 1112

[illegible]

LOCUS	AF156878	2208 bp	mRNA	ROD	16-JAN-2000
DEFINITION	Rattus norvegicus opiod growth factor receptor mRNA, complete cds.				
ACCESSION	AF156878				
VERSION	AF156878.1	GI:6693831			
KEYWORDS					
SOURCE	Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
ORGANISM	Rattus				
REFERENCE	1 (bases 1 to 2208) Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J. Cloning, sequencing, expression and function of a cDNA encoding a receptor for the opiod growth factor, [Met(5)]enkephalin				
JOURNAL	Brain Res. 849 (1-2), 147-154 (1999)				
MEDLINE	20074706				
REFERENCE	2 (bases 1 to 2208) Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J. Direct Submission Submitted (04-JUN-1999) Neuroscience and Anatomy, Penn State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA				
AUTHORS					
TITLE					
JOURNAL					
FEATURES	location/Qualifiers				
SOURCE	1..2208				
	/organism="Rattus norvegicus"				
	/strain="Sprague-Dawley"				
	/db_xref="taxon:10116"				
	/tissue_type="brain"				
	/dev_stage="18 day-old fetus"				
CDS	151..1893				
	/note="(met-5)-enkephalin binding protein"				
	/codon_start=1				
	/product="Oploid growth factor receptor"				
	/protein_id="AAF25002.1"				
	/db_xref="GI:6693832"				
	/translation="MDPPDCSTEESEDEBEGQADDTTDETGDGDDDAEARPS LFOSRTGRNWRAMQDMQRHNTPLTDODCDGDCNSFTKNEICFOPNGALIED ILNMKDNYDLLENHSYIQMLPRLRGVMMARKPTLTLEVFKSKEERLVYA YEIMLGFGYGHLEDRCGTGAIVCAQNFPFRHNLSHNNLRITRIKSLGEILEHY OAPLRFELFEETLYOHKLSPVSQALDYFLFAVCRRORRELYFAEHFPRREFWVA GNPKLRFRKPRTIPIBLTGPGCARQDKEGRSDPEAGTGGRCGSGRDI.SDGTADE DPSLNTRPFSDGCTILDNQNDKAISLPKSKRKRLTEGNROEYPGANDPQGVSEVK IALNLCEALSPLSQEPREAEPCPVARVANVEYKRKRVKGAGGDSVNSTOMQASA LPPSPCEPAOKDNGPEDNSQVGEDSKSOVPEDPNQSVELPDNSQVGEDPDP SOVGEDPNQSVQVEDPNSQVGPEDPDPNSQVVPPOAASKSPVEDPDPDVTGTSVEESE ELARIEASAEPPK"				
repeat_region	1549..1764				
	/note="imperfect"				
	/rpt_type=direct				
	/rpt_unit=1549..1575				
polya_signal	2031..2036				
polya_site	2203				
BASE COUNT	522 a 615 c 675 g 396 t				
ORIGIN					
Query Match	29.7%: Score 696.8; DB: 12; Length 2208;				
Best Local Similarity	69.9%: Pred. No. 6,7e-94;				
Matches 1045; Conservative	0; Mismatches 377; Indels 72; Gaps 5;				
Oy	48	ctcgacactccacctggaagagaacgaagatgcggaggagcgcggaacgaagagt	107		
Dd	135	cttcctgcttgacgtatgacgacgaccgacgtgcgatgccactgagagagagcga	194		
Oy	108	cgaagagcgcgaagccgcgcgcgcgcgaaggaacgcggaacgcagggagcggagc	167		
Dd	195	ggagataggcagagtagtcacgacgcgatgaaacgacacgacgacgacgacga	254		
Oy	168	ggaagagc	227		
Dd	255	cggcgacccggagagagcgcacgcccacgctctttccagtcacgagatagaggt	314		
Oy	228	ctggcagcacgaaggaacatgctgtaggtatcgcgcacaactatccgcatctgt	287		

Db 315 CTGGGCTGTAAGGAGCATACCAAGATACCGACACACCGGATTTGACAGATCA 374
 QY 288 agactgcaatgaggagcaagccaaacctgattctacagaaatgagatccgcttcctgc 347
 Db 375 AGACTGCAATGGTGCATGTGCACCTGAGCTTACAAAATGAGATCTGCTTCCAGCC 434
 QY 348 caagcgctgttccattgagacatcttcaagaactgagcagcaactatgacctctga 407
 Db 435 AAATGGGGCTCATCGAGAGACATTTTCAGAACTGAAAGCAACTATGACCTCTGGA 494
 QY 408 ggaacaatcactcacaatcagtgctgttctcttgagaaacagagatgtaactgca 467
 Db 495 AAGAAATCACTCCTACATCACTAGTGCTGCTCTCTGCGGAAACAGAGATGGA 554
 QY 468 tgcgaagcccccacagctcagtgagggtggtttaaagctcccaagagatccagga 527
 Db 555 CGCCAGACCCCTCACCTTAAGAGAGGTGAGCATTTAAAGCTCCAMGAAAGTCAGAGA 614
 QY 528 ggcgctgtcgcgggctcagagctcagctcgtggtcttcaagggatccggctgagagccg 587
 Db 615 GCGCTGTTCGGGGCTATGAGCTCATGCTGGGCTCTATGGGTTCCACCTTAGAGACCG 674
 QY 588 aggcacgagcaagtgaggccagacagaaactcagaaagcgcttcacaaactgactg 647
 Db 675 GGGCAGGGGTGTATGCGGTGTCACAGAACTTCCAGCCGCTTCCACAAATCTGAACAG 734
 QY 648 ggcgacacacaaacactccgcacacagcatcccaagctgcgtgtgactgagcct 707
 Db 735 CCAACACCCACAAACCTGCGTATTACAGCATCTCAAGTCACTGCGGAGAGTGGCT 794
 QY 708 cgaagcactcagcagcagcagctcgtcgtctctcctgagagagagcgtgtgcgcygga 767
 Db 795 AGAACTACACAGCAACCCCTGCGCTCTCTGAGAGAAACCTTGTACAGACAA 854
 QY 768 gctgcgagggtgctgcgagagtgctcctgagctactcagttgcgctgcgcgcgaca 827
 Db 855 ACTGCCAGGGTGCAGAGAGTCCCTGAGACTCTCTGCTGCTGCTGCTGCTGCTGCTG 914
 QY 828 cgaagcagcagcagctgctgactcgtcgtgagagcactccgcgcgcgcgcgcgcgcgc 887
 Db 915 CCAAGCCCGGAGCTTGTACTTGTGCTGGAGCACTTAAACCTTCCGCGAGAGTTTG 974
 QY 888 ctgaggggcccaagcaagctgagaggttcaagccagctctctgcgcactgcgtcga 947
 Db 975 CTGGGGCCCCGTGACAACTGCGAGATTCAGGCCACCAACCTTCCAGCCACTGAC 1034
 QY 948 ggcgtcagaaaggtgagagagagaaagcccgaggagcccgagcagcagcagcagc 1007
 Db 1035 GGGACCAAGGGCAGCAATAAAGATGAGGCTCCAGGGACCCCTCCCAAGAGCTGCAC 1094
 QY 1008 ccaaggtcagagcctgtgagcagagcactagcaaggtgagggcgagggg---tgagcgaag 1064
 Db 1095 CCAAGGTCGAGACTGTGATCTGGAAGGACCTGAGTGGGAGAGTGGAACAGCTGAGGA 1154
 QY 1065 gcccacagcagagcgctgagagcccaagatgagagccctctgagagggagcaggggga 1124
 Db 1155 TCCTCACTGCTGAACCAAAAGCCCTCAGATGGGGGAACCTTGATGGGAACCAAGAGGA 1214
 QY 1125 tgaagcagggggcagcaggggagatagcagggcccttaagcccaagagagcaagaa 1184
 Db 1215 TGAAG-----CTAACTCCCTGAGTCCCAAGAGAGACAGAA 1250
 QY 1185 gaagaaagctgagctgagcagcgagcagcagcagcagcagcagcagcagcagcagcag 1244
 Db 1251 AAGGAAGTTGAGGGGAACAGGAGCAGCAGTCCAGAGGAGCAGATCCCGAGGTGT 1310
 QY 1245 ctcaagagtgagaaagactcgtctgaaatttgagaggggtgctcctcagcagggcgagcct 1304
 Db 1311 CTCTGAGGTAAAGAAATTTGCCCTTAACTTGAGAGAGTGTCCCTTACCTTACAGC-- 1368
 QY 1305 cagagcagggagcccgagagtgaggcgtcaggaagcctgaggagggcagtgacgcctgcg 1364

Db 1369 -----CAGGAGCCCAAGGAGGCTGAACCCCTGTGCTC 1400
 QY 1365 ccaaccctggagcagaggtgagcagcaagtgaggaagcgaggaaggtgagtagag 1424
 Db 1401 T-----GTGGCCAGGGTGGCTATGAGTAAAGAGCGGAGAGGTGAGAGAG 1451
 QY 1425 tgcgtggagacagtgctcgtgagcagtggtggtgagcagccttgcccttgagcgc 1484
 Db 1452 GGGTGAAGGTCGATGAGTAT-----CACTAACACTCAATGACAGCCAGTCCCTGCC 1505
 QY 1485 cctgcgcccatcgaggagcccaagcgtgagacagtgagacaggggttgagga 1538
 Db 1506 TCCAAACCCCTTCAAGAGTGTCTGAGGCCCAAAAGATGGGAATGGGCCAGAGGA 1559
 RESULT 10
 AF157706/c DNA VRL 20-SEP-1999
 LOCUS AF157706 162114 bp
 DEFINITION Human herpesvirus 6B strain 229, complete genome.
 ACCESSION AF157706.1.14772.1.16947
 VERSION AF157706.1 GI:5733510
 KEYWORDS
 SOURCE
 ORGANISM
 Human herpesvirus 6B.
 Human herpesvirus 6B
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 Betaherpesvirinae; Roseolovirus.
 REFERENCE
 1 (bases 1696 to 22886)
 Pellett,P.E., Sanchez-Martinez,D., Dominguez,G., Black,J.B.,
 Anton,E., Greenmoyer,C. and Dambaugh,T.R., 1997
 A strongly immunoreactive virion protein of human herpesvirus 6
 variant B strain 229: identification and characterization of the
 gene and mapping of a variant-specific monoclonal antibody reactive
 epitope
 Virology 195 (2), 521-531 (1993)
 2 (bases 64106 to 84963)
 93331710
 REFERENCE
 Stamey,F.R., Dominguez,G., Black,J.B., Dambaugh,T.R. and
 Pellett,P.E.
 Intragenomic linear amplification of human herpesvirus 6B orlyt
 suggests acquisition of orlyt by transposition
 J. Virol. 69 (1), 589-596 (1995)
 3 (bases 64106 to 84963; 106012 to 126166)
 95074921
 REFERENCE
 Lindquester,G.J., Inoue,N., Allen,R.D., Castelli,J.M., Stamey,F.R.,
 Dambaugh,T.R., O'Brian,J.J., Danovich,R.M., Frenkel,N. and
 Pellett,P.E.
 Restriction endonuclease mapping and molecular cloning of the human
 herpesvirus 6 variant B strain 229 genome
 Arch. Virol. 141 (2), 367-379 (1996)
 4 (bases 64106 to 84963)
 96195263
 REFERENCE
 Lindquester,G.J., O'Brian,J.J., Anton,E.D., Greenmoyer,C.A.,
 Pellett,P.E. and Dambaugh,T.R.
 Genetic content of a 20.9 kb segment of human herpesvirus 6B strain
 229 spanning the homologs of human herpesvirus 6A genes U40-57 and
 containing the origin of DNA replication
 Arch. Virol. 142 (1), 103-123 (1997)
 5 (bases 106012 to 126166)
 97300856
 REFERENCE
 Lindquester,G.J., Greenmoyer,C.A., Anton,E.D., O'Brian,J.J.,
 Pellett,P.E. and Dambaugh,T.R.
 Comparison of a 20 kb region of human herpesvirus 6B with other
 human beta herpesviruses reveals conserved replication genes and
 adjacent divergent open reading frames
 Arch. Virol. 142 (1), 193-204 (1997)
 6 (bases 1 to 162114)
 97300864
 REFERENCE
 Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S., Inoue,N. and
 Pellett,P.E.
 Human herpesvirus 6B genome sequence: coding content and comparison
 with human herpesvirus 6A
 J. Virol. 73 (10), 8040-8052 (1999)
 JOURNAL
 MEDLINE
 99412318

3

[illegible]

RESULT	14
LOCUS	AF015298 13665 bp DNA VRL 10-SEP-1997
DEFINITION	Human herpesvirus 6 IE-2 protein (ie2hom) gene, complete cds.
ACCESSION	AF015298
VERSION	AF015298.1 GI:2384714
KEYWORDS	.
SOURCE	Human herpesvirus 6.
ORGANISM	Human herpesvirus 6
REFERENCE	Vituses; dsDNA viruses, no RNA stage: Herpesviridae; Betaherpesvirinae; Roseolovirus.
AUTHORS	1 (bases 1 to 13665)
TITLE	Soergel, A., Schlewe, U., Fleckenstein, B. and Neipel, F. Identification of a human herpesvirus-6 transcription unit homologous to the immediate-early 2 gene of human cytomegalovirus--evidence for an early/late gene unpublished
JOURNAL	2 (bases 1 to 13665)
AUTHORS	Soergel, A., Schlewe, U., Fleckenstein, B. and Neipel, F.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUL-1997) Institut fuer Klinische und Molekulare Virologie, Universitaet Erlangen-Nuernberg, Schlossgarten 4, Erlangen D-91054, Germany
FEATURES	Location/Qualifiers
source	1..13965
gene	/organism="Human herpesvirus 6"
cds	/strain="Uganda-1102"
	/db_xref="taxon:10368"
	1846..12117
	/gene="ie2hom"
	join(1846..1904,1993..2186,7868..12117)
	/gene="ie2hom"
	/codon_start=1
	/product="IE-2 protein"
	/protein_id="AA69858.1"
	/db_xref="GI:2384715"
	/translation="MEPAKPGGNGMSNDERMODYRPPDPMMEESIKELLESKLCDSFDLITGLSEFGLIPESNNSENVSESGDELTAEOKKGGNDVQSAKLSLCYCDADITHROPIDNRHODPIYQESRITKTDQDRIYAQTRQCTSAANIQSNBGSQVRFASLEPMLQDLPMTISNNANLNNTSLPDIYRPIYDIESTSESRNNRKYDYNFRHYPKPRGNGKYYVYDNSKYODPYRCFTRYNRGRGRHSDCSNLSADLPYENHYTPMYEDONDPKRRNNKXDFLQTNKFEESYDSMAFSTDSITQSDYNHPTKQAQITPFTKTKRHEATQNDNESTENQVLTQDVLISYRPSAKMDIKKYIVTDVPIKEALTPANGSNRDVDIOKYKKAATRCRSYOKKRRSSOTKHDENHAS

SRSDLEKRRKSNENEDKAVYIARDESKINPLISPLPLTPEPAIDFADHTDKFYSTPEP
 OIOONLHRKSTISLQDVIPISKHTPRAPTKONSYKHHSDKONPEKMSHPERTSKKN
 ITONGHQONREKYSVKVSGKATISTSPKSKTHYSHHSDGDEQKSPVKTIIOQSPZY
 CKLNPINIMDKNSAKNHTABDKNLITDINSIIRSLNLTGKPMNSKNSITMTSDNSD
 CTIDKPNKNSHKNDSPKDPNCTGKNDLHLSKOTYMAKTRAPSPASRRASRSDSRAS
 SRASRRASRSDSRASRRASRSDSRASRRASRRASRRASRRASRRASRRASRRAS
 ASSRRASRRASRSDSRASRRASRRASRRASRRASRRASRRASRRASRRASRRAS
 SRDSRRASRSDSRSDSRASRRASRRASRRASRRASRRASRRASRRASRRASRRAS
 ASSRRASRRASRSDSRASRRASRRASRRASRRASRRASRRASRRASRRASRRAS
 SRASRRLOIYICDSNKRQTPHDTSLNTPELSEILFRCGEDINFKNPAIRLQCPN
 HNDQYNDPRERPHAIRMKRSSESTINDDSMSMRCSHCNRPDSLTIVARRKKHKC
 SSSISSISSIEENCRNSHIVGKEKTFPYIOSSTRSSSSSSSSSALSOSKTKKKG
 RKIYQKNDKOIKRSKSDKHKTTDMSSDYSEENRHADVKNPEAEGERPLHSSPFINH
 EOSHNSENAIIDEOBKRAPIHTTSHLOQOVNPLHNTCKKCKRRPDDSDOSIINPCK
 KRISGAQOTSESEVDPDLCYRDVRLKERSYKREPIHGGRAVATKFOKLFRTMTA
 PEYQOIKRCKPCKNEKNTKEAYVYDCCNLSNNAALIMVFTSKSYKQIOTKOKLMT
 PNITVSEAFKKNNAHPKYIYDDEKRRITQLOQOGPOBLMDKFEENTHTHDLFSRSDVK
 TMLIYAAPIDFVGAVTCKNKYAKDNKRETELVRCSITIIDDPNISTINPISKERKSKF
 STJLSC"

Query Match	4.78;	Score 110.6;	DB 72;	Length 13965;
Best Local Similarity	54.2%;	Pred. No. 7.2e-08;		
Matches 224; Conservative	0;	Mismatches 189;	Indels 0;	Gaps 0;

QY	1585	ccatcgagagccccagccccagccccagcagagactgtgaaggagccagccagccagagc	1644
Db	9845	CCAGCAGAGGCTTCACAGAGGATTTCTAGCAGAGCTTCACAGAGGAGCTTCACAGAGGGATT	9904
QY	1645	ccatcgagagccccagccccagccccagcagagactgtgaaggagccagccagccagagc	1704
Db	9905	CTAGCAGAGGCTTCAGATTAAGGCTTCACAGAGGCTTCAGCAAAGCTTCACAGAGAGCTT	9964
QY	1705	ccatcgagagccccagagcctccgcccgcgcgcagagactgtgaaggagccagccagagacc	1764
Db	9965	CCAGCAGAGGCTTCACAGAGGCTTCACAGAGGATTTCTAGCAGAGGCTTCAGCAAAGCTT	10024
QY	1765	ccatcgagagccccagccccagccccagcagagactaaaggagtgtgcccagcccgagagc	1824
Db	10025	CCAGCAGAGGCTTCACAGAGGATTTCTACAGAGGCTTCAGTAGAGGCTTCACAGAGGGCTT	10084
QY	1825	ccatcgagagccccagagccccgcgcgcgcgcgcagagactgtgaaggagccagccagagagc	1884
Db	10085	CAGCAAAGGCTTCAGACAGAGCTTCACAGAGCTTCAGAGAGGCTTCACAGAGGGATT	10144
QY	1885	ccatcgagagccccagagccccgcgcgcgcgcgcagagactgtgaaggagccagccagagagc	1944
Db	10145	CTAGCAGAGGCTTCAGCAAAGGCTTCACAGAGGCTTCACAGAGGATTTCTAGAGAGCTT	10204
QY	1945	ccatcgagagccccagagccccgcgcgcgcgcgcagagactaaaggagtgtgcccagc	1997
Db	10205	CCAGCAGAGGATTTCTACAGAGGATTTCTACAGAGGCTTCACAGAGAGGCTTCAGC	10257

RESULT	15		
LOCUS	HHU13194/c		
DEFINITION	HHU13194	21191 bp	DNA
	Human herpesvirus 6 replication origin-binding protein (HDRFO), partial cds, helicase-primase component (HDRFL), vitron protein (HDRFL), putative helicase (HDRF2), putative phosphoprotein (HDRFL), replication protein (EDRF2), putative uracil-DNA glycosylase (EDRF4), putative gl. glycoprotein (EDLF3), putative glycoprotein (EDLF2) genes, complete cds, and major immediate-early protein 2 (BCLF0) gene, partial cds.		VF1
ACCESSION	U13194		17-JUN-1995
VERSION	U13194.1	GI:862477	
KEYWORDS			
SOURCE	Human herpesvirus 6.		
ORGANISM	Human herpesvirus 6		
	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae; Roseolovirus.		

[illegible]

CDS

/evidence-not_experimental
complement(10022..10351)
/gene="EDLF5"
/codon_start=1
/product="unknown"
/protein_id="AAA68469.1"
/db_xref="GI:662101"
/translation="NMMLAYFLYCRQLLAVLVLEPPRLCGNTLREYSIHLPLSTCH
ISDSFYLSRTSNTLFLCHLYCHOKSHOKSHOKTKTGLKLSNTIENIKMFNSSP
YTLSTYA"
11477..12652
/gene="EDRF1"
11477..12511
/gene="EDRF1"
/note="putative phosphoprotein; similar to HH6 CB7R,
GenBank Accession Number L14772 and HCMV UL112, Swiss-Prot
Accession Number P16768"
/codon_start=1
/protein_id="AAA68470.1"
/db_xref="GI:662102"
/translation="WAEERAYGSFDDVMEYVQOIVTESIQIKRLHFGSGCLIEFLGD
SGTCTFCGGWISMTSDTNSMGLTVDIGDKYTNARGFLCSKITSISO
NTEGRDLITVSQENNKLOITEVTLTKYKEHIDINLDPCKIEKEKCALDRKH
DDEHRRSGKOKERKVEDDKKDDRRKROBERKRNEDKOPDKKESDGRKCIIV
VNFYSCVTYVDFEFLNSICFVAISLPKPKRKQYHDMKRNLEOSHEDGITLSTIV
NGAVGALPQCISINHEDDQHDELDRAYADVGVSQSPKRRSSGLTCYKADLFI
LGDEPRKSIQ"
12483..13180
/gene="EDRF2"
<12483..13079
/gene="EDRF2"
/note="putative phosphoprotein; similar to HCMV UL113,
Swiss-Prot Accession Number P17151"

gene
CDS

gene
CDS

Query Match
Best Local Similarity 50.8%; Pred. No. 4.7e-06;
Matches 234; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 1543 acagaaagtcgaagcggcccaagaaggtaccctcggagaccatcgagaaccacgagc 1602
DB 19520 AAGATTGACCAAGCTTCTTAAAGGGCTCCAGCAGAGCTCAAGCAGAGCCTCCAGCA 19461
QY 1603 cccagcccgagacccgagcggagcgaagccagcccgagacccatcgagaccacgagc 1662
DB 19460 GGGATTCTAGCAGGGGCTCAAGCAGAGCCTCCAGCAGAGCCTCCAGCAGGATTTAGCA 19401
QY 1663 ccccgccagcagacccgagcggagcgaagccagcccgagacccatcgagaccacgagc 1722
DB 19400 GAGCTTCCAGCAGAGCCTCCAGCAGGATTTAGCAGAGCCTCCAGTAGAGCCTCCAGCA 19341
QY 1723 ctcgcccggcagcagcctcgagcggagcgaagcccgagaccccatcgagaccacgagc 1782
DB 19340 GGGCTTCAAGCAAGAGCTTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCA 19281
QY 1783 cccagcccgagacccgagcggagcgaagcccgagacccatcgagaccacgagc 1842
DB 19280 GGGATTCTAGCAGGGGCTCAAGCAGAGCCTCCAGCAGAGCCTCCAGCAGGATTTAGCA 19221
QY 1843 ccccgcccgagcagcctcgagcggagcgaagcccgagacccatcgagaccacgagc 1902
DB 19220 GAGCTTCCAGCAGGATTTAGCAGAGGCTCCAGCAGAGCCTCCAGCA 19161
QY 1903 ccccgccggcagcagcctcgagcggagcgaagcccgagacccatcgagaccacgagc 1962
DB 19160 GGGATTCTAGCAGGGGCTCAAGCAGAGCCTCCAGCAGAGCCTCCAGCA 19101
QY 1963 cccagcccgagcagcctcgagcggagcgaagcccgagaccacgagc 2003

DB 19100 GGGCTTCCAGCAGGGGCTCAAGCAGAGCCTCCAGTAAAGCC 19060

Search completed: October 12, 2000, 18:33:21
Job time: 36977 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2000, 18:33:21 ; Search time 17897.3 seconds
(without alignments)
558.597 Million cell updates/sec

Title: US-09-431-843b-11
Perfect score: 2289
Sequence: 1 tagaattcagcgccgcctga.....cctctcgagtgaggcgcg 2289

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_p11:*
8: gb_p12:*
9: gb_p17:*
10: gb_p18:*
11: gb_p19:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_p1:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_p13:*
34: gb_p14:*
35: gb_p15:*
36: gb_p19:*
37: gb_p19:*
38: gb_p19:*
39: gb_p19:*
40: gb_p19:*
41: gb_p19:*
42: gb_p19:*
43: gb_p19:*

44: gb_p19:*
45: gb_p19:*
46: gb_p19:*
47: gb_p19:*
48: gb_p19:*
49: gb_p19:*
50: gb_p19:*
51: gb_p19:*
52: gb_p19:*
53: gb_p19:*
54: gb_p19:*
55: gb_p19:*
56: gb_p19:*
57: gb_p19:*
58: gb_p19:*
59: gb_p19:*
60: gb_p19:*
61: gb_p19:*
62: gb_p19:*
63: gb_p19:*
64: gb_p19:*
65: gb_p19:*
66: gb_p19:*
67: gb_p19:*
68: gb_p19:*
69: gb_p19:*
70: gb_p19:*
71: gb_p19:*
72: gb_p19:*
73: gb_p19:*
74: gb_p19:*
75: gb_p19:*
76: gb_p19:*
77: gb_p19:*
78: gb_p19:*
79: gb_p19:*
80: gb_p19:*
81: gb_p19:*
82: gb_p19:*
83: gb_p19:*
84: gb_p19:*
85: gb_p19:*
86: gb_p19:*
87: gb_p19:*
88: gb_p19:*
89: gb_p19:*
90: gb_p19:*
91: gb_p19:*
92: gb_p19:*
93: gb_p19:*
94: gb_p19:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2288.6	100.0	2363	11 AF172452	AF172452 Homo sapi
2	2212.6	96.7	2423	11 AF172451	AF172451 Homo sapi
3	2149.4	93.9	2483	11 AF172453	AF172453 Homo sapi
4	2130.8	93.1	2423	11 AF109134	AF109134 Homo sapi
5	1567.8	68.5	160241	35 HS885L7	AL035669 Human DNA
6	1555	67.9	6670	11 AF112980	AF112980 Homo sapi
7	1351	59.0	1676	11 AF172450	AF172450 Homo sapi
8	977.4	42.7	1232	11 AF172449	AF172449 Homo sapi
9	698.4	30.5	2208	12 AF156878	AF156878 Rattus no
10	106.4	4.6	30752	72 HHU92288	U92288 Human herpe
11	103.2	4.5	161573	72 AB021506	AB021506 Human her
12	101.4	4.4	162114	72 AF157706	AF157706 Human her

13	92.2	4.0	4742	72	AF015297	AF015297 Human her	
14	92.2	4.0	13965	72	AF015298	AF015298 Human her	
C	15	83.6	3.7	21191	72	HHU13194	HHU13194 Human herpe
C	16	80.4	3.5	159321	72	HHV6AGNM	X83413 Human herpe
C	17	79.	3.5	167830	43	AC020678	AC020678 Homo sapi
C	18	77.6	3.4	303819	30	AE003509	AE003509 Drosophil
C	19	76	3.3	7218	5	166494	166494 Sequence 14
C	20	75.8	3.3	1072	73	HS21EA	M29385 Herpes simp
C	21	75.8	3.3	6192	5	AR030811	AR030811 Sequence
C	22	75.8	3.3	6449	5	AR030812	AR030812 Sequence
C	23	75	3.3	328621	47	AC026285	AC026285 Homo sapi
C	24	73.8	3.2	3941	12	RNLAR2	X83546 R.novglicu
C	25	73.4	3.2	193453	45	AC033920	AC033920 Homo sapi
C	26	73	3.2	219455	39	AC012179	AC012179 Homo sapi
C	27	72.2	3.2	1177	32	PCPEPT	X74358 P.carnae Po
C	28	72	3.1	120469	52	AC074307	AC074307 Mus muscu
C	29	70.4	3.1	135999	66	AC078925	AC078925 Homo sapi
C	30	70.4	3.1	163624	52	AC073862	AC073862 Homo sapi
C	31	70.4	3.1	252887	52	AC073527	AC073527 Homo sapi
C	32	70.2	3.1	166835	37	AC009573	AC009573 Homo sapi
C	33	70	3.1	1227675	49	AC044845	AC044845 Mus muscu
C	34	69.8	3.0	2479	12	RNBS7960	U87960 Rattus norv
C	35	69	3.0	759	32	ENHR2NPX	M75950 Entamoeba h
C	36	66.6	3.0	2423	11	AF109134	AF109134 Homo sapi
C	37	66.6	3.0	6670	11	AF112980	AF112980 Homo sapi
C	38	66.6	3.0	160241	35	HS8857	AL035669 Human DNA
C	39	68.6	3.0	218735	39	AC011899	AC011899 Homo sapi
C	40	68.4	3.0	2026	31	BBOEW77K	M95675 Babesia bov
C	41	68	3.0	178795	68	AL355677	AL355677 Homo sapi
C	42	67.2	2.9	1150	73	HS4LIR13	J02079 epstein-bar
C	43	67.2	2.9	5452	13	U02454	U02454 Cloning vec
C	44	67.2	2.9	9600	5	A92865	A92865 Sequence 1
C	45	67.2	2.9	10596	5	I25041	I25041 Sequence 15

[illegible]

```

RESULT      1
AF172452
LOCUS       AF172452      2363 bp      mRNA      PRI      19-APR-2000
DEFINITION Homo sapiens opioid growth factor receptor mRNA, complete cds.
ACCESSION  AF172452
VERSION     AF172452.1
KEYWORDS   GI:7595304
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2363)
            Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
            Cloning, sequencing, chromosomal location, and function of cDNAs
            encoding an opioid growth factor receptor (OGFR) in humans
            Brain Res. 856 (1-2), 75-83 (2000)
JOURNAL    2 (bases 1 to 2363)
            Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
            Direct Submmission
            Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
            Pennsylvania State University College of Medicine, 500 University
            Drive, Hershey, PA 17033, USA
FEATURES   location/Qualifiers
            1..2363
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="20"
            /map="20q13.3"
            /issue_type="placenta"
            34..2007
            /note="[Met-5]-enkephalin binding protein; OGFR;
            alternatively spliced"
            /codon_start=1
            /product="opioid growth factor receptor"

```

QY	241	aggcaatggttaagttatcgacacaacatctcggaatctgtgtggaaagagatctgcaatgg	300
Db	241	AGGGCAATGTTAGTATTCGGCAACAATCTCGGATCTGGTGGAAAGAGATCGAATGGG	300
QY	301	gacacgccaacacctgagttttctacagaatatagatccgcttcctccgccaacgctgtttc	360
Db	301	GACACGCCAACCTGAGTTTCTACAGAAATATGATTCGGCTTCCTGGCCAAAGGCTGTTTC	360
QY	361	attagagacattctcagaatgtagcggacaacatatacctcctcttggaggaaatcacttc	420
Db	361	ATTAGAGACATTCTTCGAACTGGACGACCACTATTAACCTCTTGGAGAAATCACTCC	420
QY	421	tacatccagttgctgtttcctctctcagaaacacagagatgaaatgcaatgccaaagccctc	480
Db	421	TACATCCAGTGTGGTGTTCCTCTCTCGAAGAACAGAGTGAATCGCATGCCAACCCCTC	480
QY	481	acgcctcaggagagtgctccggagtggtttaaaagcttcccaagagatccagaagcggctgtccgg	540
Db	481	ACGCTCAAGGGAGGTCTGAGAGTGTAAAAAGCTCCAGGAGATCCAGGAGCGGCTGTGTCGG	540
QY	541	gcctacagagctcaatgctcgggctctctacaggagatccggctctggagagacagagcaggcaq	600
Db	541	GCTACAGAGCTCAATGCTAGGGCTTCTACGGGATCCGGTGGAGGAAACGAGACGGGCACG	600
QY	601	gtggagccgagacagaagaactacagaaagcgcttcagaaacctgaaactgacactgacgacacaac	660
Db	601	GTTGGCCGAGACACGAACCTACCAAAACCGCTTCCAGAACTCGAAGCTGGCGAGCCAAAC	660
QY	661	aaactccgacatcaacagcatcctcaagtgcggtgtagctgagacccctcggagcacttcag	720
Db	661	AACTCCGCATCAACACCATCTCAATCCGCGTGTAGCTGAGACCTCGAGCACTTCAG	720
QY	721	ggcgcaactggtcggctctctctctcctggaggagagacgtgtgtcggcgggagcgtgcgcgggtg	780
Db	721	GGCGCACTGTGTCCTCTTCTCTGAGAGACGCTGTGGCGCGGAGAGCTCCGCGGGTG	780

QY 781 cgcagagatgcccctggaactcatgtctgcgcgtgcctccgagaccagcgcccgag 840
|||||
Db 781 CGCGAGAGTGCCTTGATCTACTTTCATGTTCCGCGCTGCTGCGACACACGCGCCGAG 840
QY 841 ctgtgtcacttcgccttggagacacttcggcccgctgcaagttcgtctggggcccaa 900
|||||
Db 841 CTGTGACACTTCGCTGGAGACACTTCGGCCCGCTGCAAGTTCGTCTGGGGGCCCAA 900
QY 901 gacaagctgcggaggttcaagcccagctctctgcccacgcgtcgagggctcagaag 960
|||||
Db 901 GACAGAGTCGGAGGTTAAGCCCACTCTGCGCCATCCGCTCGAAGGCTCCAGGAG 960
QY 961 gtgag 1020
|||||
Db 961 GTGAG 1020
QY 1021 tctgtgag 1080
|||||
Db 1021 TGTGGGCCAGAGCATAGCAAGGTTGGGGGCAAGGTTGAGAGAGGGGCCACAGCAG 1080
QY 1081 gtgag 1140
|||||
Db 1081 GTGAG 1140
QY 1141 gggag 1200
|||||
Db 1141 GGGAG 1200
QY 1201 agcgag 1260
|||||
Db 1201 AGCGGAG 1260
QY 1261 atcgcctgaatttggagagagagagagagagagagagagagagagagagagag 1320
|||||
Db 1261 ATCGCTCTGAATTTTGAAGGGGTGTGCTCTCAAGCCAGGCGACGCTCAGAGAGAGAG 1320
QY 1321 gaagtgag 1380
|||||
Db 1321 GAAGTGAG 1380
QY 1381 aggtgtgag 1440
|||||
Db 1381 AGGGTGAG 1440
QY 1441 gcggtgtgag 1500
|||||
Db 1441 GCGGTGTGAG 1500
QY 1501 caccgag 1560
|||||
Db 1501 CACCCGAG 1560
QY 1561 cccag 1620
|||||
Db 1561 CCCAAG 1620
QY 1621 gcaagtgag 1680
|||||
Db 1621 GCAGGAG 1680
QY 1681 gcaagtgag 1740
|||||
Db 1681 GCAGGAG 1740
QY 1741 acaagtgag 1800
|||||
Db 1741 ACAAG 1800
QY 1801 gcaagtgag 1860
|||||
Db 1801 GCAGGAG 1860

QY 1861 gcaagtgag 1920
|||||
Db 1861 GCAGGAG 1920
QY 1921 acaagtgag 1980
|||||
Db 1921 ACAAG 1980
QY 1981 tcttctgcaag 2040
|||||
Db 1981 TCTTCTGCAAGAGTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 tctcctgtcag 2100
|||||
Db 2041 TGTCTCTCTCAG 2100
QY 2101 tctgtacccgtgacccatgacacagagagagagagagagagagagagagagagag 2160
|||||
Db 2101 TGTGTACCCGTGACCCATGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 ccag 2220
|||||
Db 2161 CCAG 2220
QY 2221 tcttctcccaag 2280
|||||
Db 2221 TGTCTTCCCAAG 2280
QY 2281 ggggag 2340
|||||
Db 2281 GGGGAG 2340

RESULT 2
AF172451 2423 bp mRNA PRI 19-APR-2000
LOCUS Homo sapiens oploid growth factor receptor mRNA, complete cds.
DEFINITION AF172451
ACCESSION AF172451.1 GI:7595302
KEYWORDS
SOURCE
ORGANISM human.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2423)
Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
Cloning, sequencing, chromosomal location, and function of cDNAs
encoding an oploid growth factor receptor (OGFr) in humans
JOURNAL Brain Res. 856 (1-2), 75-83 (2000)
MEDLINE 20143540
2 (bases 1 to 2423)
Zagon,I.S., Verderame,M.F., Allen,S.S. and McLaughlin,P.J.
Direct Submission
Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA
FEATURES
source
1. .2423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue-type="placenta"
/note="Met-5]-enkephalin binding protein; OGFr;
alternatively spliced"
/codon_start=1
/product="oploid growth factor receptor"
/protein_id="AA64404.1"
/db_xref="GI:7595303"
/translation="MDPDCDSWEDEDEDAEDCEDEGEAGARADAGADEDE
SEEPRAIPSSFOSRMTGSNMRATRMCRVRYHYPTLVKDCDGDPTNLSFYRNEIR
FLPNCFTIEDILQWMTDNYDLLEDNHSYIQWLPFLRBPVNMHAKPLTLREVEYFKSS

OEIIEVLVRAVELMIGFYGILIEDRGSTGYVARAONYKRPONLWBSHNNILRIRLIK
 SCPELSLEHPOLYREFLEETLYIRRELPGYRQALNOYKXPRNOLVHRODLYHAFME
 HERRRCKRWGPOKILRRKRSSTPHLESGRKUEBGSBDDPHENASTOGRTGPHN
 SKGGKRDDEGQPSRVSERFODAGPFRLESGDGDAGHGEDEBPPLPSKSKKRLDLSTR
 EOPTEBPPOSASVEKRTALNLEGALSOGSLRTIGTIDVGODGEAVOPOROLGAKR
 VADYVRRKRRKUEDEGASDAVAASGAQTLTLAGSPABSGKAAHSENGVEDEEGRT
 GPKCTGPGSPSETGPPSPAGAGDEPAFSEETGPRPAGAGAGDEPSPSETGLLP
 AGGAPGDEPAELPSETTGPSPAGPTRDEPAEESPSTGPRRAGPAGDPAEESPSTPBP
 RPAPAGDEPPEPSESETTGPSPAGPTRDEPAEAKGAEALQDPAEYESSAKSGKP"

503 a 754 c 843 g 323 t

BASE COUNT
 ORIGIN

Query Match	96.78;	Score 2212.6;	DB 11;	Length 2423;
Best Local Similarity	97.38;	Pred. No. 0;		
Matches 2285; Conservative	0;	Mismatches 4;	Indels 60;	Gaps 1;

[illegible]

QY	841	ctggtgcaacttcgccccttggtgagcaacttcgcgccccgctgtcaagtttgctctggtggccccca	900
Db	841	ctggtgcaacttcgccccttggtgagcaacttcgcgccccgctgtcaagtttgctctggtggccccca	900
QY	901	gacaagctgcgtgagtttccaagccccagctctctgtccccatccgctctgtagaggtcccaagag	960
Db	901	gacaagctgcgtgagtttccaagccccagctctctgtccccatccgctctgtagaggtcccaagag	960
QY	961	gttgagagagaaagaaagcccccttggtgagccccgcacccagagagccagcaccccaggtctgagac	1020
Db	961	gttgagagagagaaagaaagcccccttggtgagccccgcacccagagagccagcaccccaggtctgagac	1020
QY	1021	tgttggtccaagagcatbaacaaagtgtggtgggtcagagtgtgacagaggtggccccacgcagagac	1080
Db	1021	tgttggtccaagagcatbaacaaagtgtggtgggtcagagtgtgacagaggtggccccacgcagagac	1080
QY	1081	gttgagagcccccaagatltggtgggaaccccttgtagaagagcccaagggatgtaggtcaagggagac	1140
Db	1081	gttgagagcccccaagatltggtgggaaccccttgtagaagagcccaagggatgtaggtcaagggagac	1140
QY	1141	gggtgagagataaggtccgtgagacccttcaagcccccaagagagcaagaagagagagctgtgagctg	1200
Db	1141	gggtgagagataaggtccgtgagacccttcaagcccccaagagagcaagaagagagagctgtgagctg	1200
QY	1201	agccgtgcgtgagcagccgcccccaagagccagccctcaagatgtcctcagaagtgtgagag	1260
Db	1201	agccgtgcgtgagcagccgcccccaagagccagccctcaagatgtcctcagaagtgtgagag	1260
QY	1261	atcgctctcgaatttgtaggggtgtgtgcacctagcccaagagagagctcaagagccgtgagagacag	1320
Db	1261	atcgctctcgaatttgtaggggtgtgtgcacctagcccaagagagagctcaagagccgtgagagacag	1320
QY	1321	gaagtggtgcgtgtcagagacccttggtgtagcagctgtcagccctgcgcgcaaaaccccttgtagac	1380
Db	1321	gaagtggtgcgtgtcagagacccttggtgtagcagctgtcagccctgcgcgcaaaaccccttgtagac	1380
QY	1381	aggtgtgcgcagcaaggtgtagagaaacgtgagagaggtgatgtagagtgtctgtggtgagagctgtc	1440
Db	1381	aggtgtgcgcagcaaggtgtagagaaacgtgagagaggtgatgtagagtgtctgtggtgagagctgtc	1440
QY	1441	gggtgtggtcaggttgtgtgtgccccagaccttgcgaccttgccgtgtcccttgcacctcaggg	1500
Db	1441	gggtgtggtcaggttgtgtgtgccccagaccttgcgaccttgccgtgtcccttgcacctcaggg	1500
QY	1501	caaccacaaggtctgtagacaacagtgtgaaagaggtgtgtgtagagagacacaagaagtctgaaacggg	1560
Db	1501	caaccacaaggtctgtagacaacagtgtgaaagaggtgtgtgtagagagacacaagaagtctgaaacggg	1560
QY	1561	cccaagaagaagtataccctctgtgtagcccatctgtagagcccccaagcccccaagcagacagacct	1620
Db	1561	cccaagaagaagtataccctctgtgtagcccatctgtagagcccccaagcccccaagcagacagacct	1620
QY	1621	gcaaggtgagccaggtccaggtccgtgagagcccatctgtagagcccccaagcccccaagcagacagacct	1680
Db	1621	gcaaggtgagccaggtccaggtccgtgagagcccatctgtagagcccccaagcccccaagcagacagacct	1680
QY	1681	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1740
Db	1681	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1740
QY	1741	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1800
Db	1741	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1800
QY	1801	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1860
Db	1801	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1860
QY	1861	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1920
Db	1861	gcaaggtgagccaggtccaggtccgtgagagcccccaagcccccaagcagcccccaagcagacagacct	1920

||||| 1921 GCAAGGACCAAGCAGGAGCCATCGAGAGCCAGGCCAGCCAGCCAGGAGAGCT 1980
QY 1921 acaagagatgagccaagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1980
Db 1981 ACAAGGATGAGCCAGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
QY 1981 tctctcgaagtcgtggaagccttaaggaagagtgccgtcgagctgtgtcc 2040
Db 2041 TCTTCTGCAAGTGTGGGAGCCTTAAGAAAGAGTCCCTGGGCTCTTGGTCTCC 2100
QY 2041 tctcctctgcaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 2100
Db 2101 TGTCCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
QY 2101 tctgagcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 2160
Db 2161 TCGTGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
QY 2161 ccaagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 2220
Db 2221 CCAAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
QY 2221 tctctcgaagtcgtggaagccttaaggaagagtgccgtcgagctgtgtcc 2280
Db 2281 TGTCCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
QY 2281 gggggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 2340
Db 2341 GGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG

RESULT 3

AF172453 2483 bp mRNA PRI 19-APR-2000
LOCUS AF172453 Homo sapiens oploid growth factor receptor mRNA, complete cds.
DEFINITION AF172453.1 GI:7595306
ACCESSION AF172453.1
VERSION AF172453.1
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 2483)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Cloning, sequencing, chromosomal location, and function of CDNA
encoding an oploid growth factor receptor (OGFR) in humans
Brain Res. 856 (1-2), 75-83 (2000)
JOURNAL MEDLINE
20143540
2 (bases 1 to 2483)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Direct Submission
Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA
JOURNAL location/Qualifiers

FEATURES

source
1. 2483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue_type="placenta"
34. 2127
/note="[Met-5]-enkephalin binding protein; OGFR;
alternatively spliced"
/codon_start=1
/product="Oploid growth factor receptor"
/protein_id="AA64406.1"
/db_xref="GI:7595307"
/translation="MDPPDCSTWEDEDAEDDEDCEGAAGARADADDEDE
SEEPRAAPSSFOSRMTGKRWKATKRCRKHRYTPULVEDCGDTPNLSFRNEIR
FLPVGCIEDILQNTDNYDLLEDNHSYIOWLPFLRPGVNMWAKPLTLREVEYFKS
OEIOERLVRAVELMAGFYGIRLEDRTGTGVRANQYKRFQENWNRSHNMLRITRIIK

SPCELSLEHFOAPLVRFLEETLVREELPGVQSAIDYEMFVRCROROLVHAME
HFRPCKFTWGPDPDLKREKPSPIPHLESGKVEEGSPGPDHASTQGRCEBEH
SKGGVVDGPODROPSVPEODAGPLRSDPDGKHEGDEPREPLSPKRIEISLR
EOPPTPGQSAEVEKILNINEGCLSSGSLRTGQOEVGGDDPGANOPCQPIGAR
VADKVRKRNVDDGTGDSAAVAGQTLALGSPAFSPASHPAHSENVSEDTBERT
GPREKTPGSPSETPGSPAGDEPAKTPSETPGSPAGPTRDEPAESPSPTPGP
ACPADPEPAESPSPTPGSPAGDEPAKTPSETPGSPAGPTRDEPAESPSPTPGP
RPAAGDEPAESPSPTPGSPAGDEPAKTPSETPGSPAGPTRDEPAESPSPTPGP
LDAEVESSAKSKP"
BASE COUNT 518 a 778 c 861 g 326 t
ORIGIN

Query Match 93.9%; Score 2149.4; DB 11; Length 2483;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 2283; Conservative 0; Mismatches 6; Indels 120; Gaps 1;

QY 1 tagaattcaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 60
Db 1 TAGAATTCAGAGGCGCGCTGATTTCTACCGAGCATGAGACGCCGACTGCACTCCACC 60
QY 61 tggagagagagagagagagagatgagagagcgagagagagagagagagagagagag 120
Db 61 TGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 121 ggc 180
Db 121 GCC 180
QY 181 gc 240
Db 181 GCG 240
QY 241 agggacatgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 300
Db 241 AGGGACATGTTAGTATGCGCAACTATCCGATCTGTGGAGACGACTGCATAATGGG 300
QY 301 gacacgcacaacatgagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 360
Db 301 GACACGCCCAAACTGAGTTTCTACAGAAATGAGATCCGCTTCCGCCCAAGCGCTGTTTC 360
QY 361 attgagacatctcttcgaagtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 420
Db 361 ATTGAGGACATTTCTGAGAACTGAGCGAGCAACTATGACCTCTTGGAGCAATCACTCC 420
QY 421 taactcaagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 480
Db 421 TACATCCAGTGGCTGTTCTCTCTCGAAGAACAGAGTGAATGCAATGCCAACCCCTC 480
QY 481 acgctcagagagagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 540
Db 481 ACGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 541 gcttacagagtcagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 600
Db 541 GCTTACGAGGCTCAGTGTGGGCTTCTACGGGATCCGGCTGGAGGAGCGAGGAGCGGACG 600
QY 601 gtgggc 660
Db 601 GTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
QY 661 aactcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 720
Db 661 AACTCTCGCATCACAGCATCTCAAGTCCGCTGTAGAGCTGAGAGCTCGAGCACTTCCAG 720
QY 721 gcgcacatgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 780
Db 721 GCGGACAGTGTGCGCTTCTCTGAGAGGAGCGCTGGGCGGCGGAGGAGGAGGAGGAGGAG 780
QY 781 cgc 840
Db 781 CCGGAGAGTGGCTCGGACTACTTATGTTCGCTGCGCTGCGGAGGAGGAGGAGGAGGAGGAG 840

```

OY 841 ctgtgtcaactctcgcttggaggaacttccggtcccgctgcaagtctgttggggcccaa 900
      |||
Db 841 CTGTGTCACTTCGCTTGGAGACATTCGCGCCCGCTGCAAGTTGTTGGGGCCCAA 900
OY 901 gacaagctgaggaagttaaaagcccaagctctctgtcccaatccgctcgaaggttccaagaa 960
      |||
Db 901 GACAAAGCTGCGGAAGTTAAAGCCCAAGCTCTGTGCGCATTCGGCTGAGGGGTCCAGAGAG 960
OY 961 gtcgaggaaggaagaaagcccggtggagcccgacacagagccagagcccaaggttcgaac 1020
      |||
Db 961 GTGAGAGAGGAAGAAAGACCCCGGGGACCCGACACAGAGCCAGACACCAAGGTTGAGACC 1020
OY 1021 tctgtggccagagcatagcaagaggttggggcaggttgcgcaggggggcccccaacagagac 1080
      |||
Db 1021 TGTGAGCCAGACATAGCAAGAGGTGGGGCAGGGGTGAGCGAGGGGCCCCACACAGGAGC 1080
OY 1081 gtcgagcccgaggaatgagggagcccttggagagagccaggggaatgagagcaggggacac 1140
      |||
Db 1081 GTGAGCCCGAGATGCGGGACCTTGGAGAGAGCCAGGGGATGAGGCGAGGGGCCAC 1140
OY 1141 ggggaagaataagccgagaccccttaagccccaagagacaaagaaaggaagctgagctg 1200
      |||
Db 1141 GGGGAAGATAGGCCCGGAGCCCTTAAGCCCCCAAGAGACAAAGAAAGAGAACTGAGACTG 1200
OY 1201 agcgggagagcagcgcgccccacagagcccgccctcagagtgtcctcagaggttgaagaa 1260
      |||
Db 1201 AGCCGGCGGAGACAGCCGCCCAAGGGCCAGGCCCTCAGAGTGCCTCAGAGTGGAGAG 1260
OY 1261 atgcgtcgaatttggagaggtgtgcccctcagccagagccagcctaagagacggggacag 1320
      |||
Db 1261 ATGCTCTGAATTTTGGAGGGGTGTGCTCCTCAAGCCAGGGCATGAGGAGGGGAGCCAG 1320
OY 1321 gaagtgggcgttcaagaccccttggggaggaagtgacgtgcagccctgcgcgaacccctggagcc 1380
      |||
Db 1321 GAAGTGGCGGTCAGAGACCTTGGGAGGAGCATGCAAGCCCTGCCGCAACCCCTTGAGAGCC 1380
OY 1381 aggtgtgcccgaagaagtgaaggaagggagaggtgaggtgaggtgtcggggagcaagtgt 1440
      |||
Db 1381 AGGGTGGCCGAGAAAGTGAAGAGCGAGAGAAAGTGAAGGTACTGGGAGACGTGCT 1440
OY 1441 ggcgtgagcagtggtgtgcccccaagccttgcgttgcgggttcccggtcccatcgggg 1500
      |||
Db 1441 GCGGTGGCCAGTGTGTGTCGCCAGACTTGGCCTTGCCGGGTCCCGTCCCATCGGG 1500
OY 1501 caaccccaaggtctgacaacagtgaagaacgggtgtgagagagacacagaaggttgaacggg 1560
      |||
Db 1501 CACCCCAAGGCTGGACACAGTGAAGACGGGTGTGAGGAGACACAGAAAGTGAACGGGG 1560
OY 1561 cccaaagaggtaccccttggagcccatctggaagaccccgagcccgagccca----- 1611
      |||
Db 1561 CCCAAAGAAAGTACCCCTGGAGCCCATCTGAGAACCCCAAGCCCCCAAGCCACAGAGACT 1620
OY 1612 ----- 1611
      |||
Db 1612 GCAAGGGAGAGACCAAGCAAGACCCCATCGAGAACCCCAAGCCCCCGCCGAGAGACT 1740
OY 1621 gcaagggagagcagccagccagagcccatctggaagaccccgagcccgcccgcaagacct 1680
      |||
Db 1621 GCAAGGGAGAGACCAAGCAAGACCCCATCGAGAACCCCAAGCCCCCGCCGAGAGACT 1800
OY 1681 gcaagggagagcagccagccagagcccatctggaagaccccgagcccgcccgcaagacct 1740
      |||
Db 1681 GCAAGGGAGAGACCAAGCAAGACCCCATCGAGAACCCCAAGCCCCCGGAGAGACT 1860
OY 1741 acaagggatgagccagccagagcccatctggaagaccccgagcccgcccgcaagacct 1800
      |||
Db 1741 ACAAGGGATGAGCCCAAGAGCCCATCTGGAGAACCCCAAGCCCCCGCCGAGAGACT 1920
OY 1861 gcaagggagagcagccagccagagcccatctggaagaccccgagcccgcccgcaagacct 1860
      |||
Db 1861 gcaagggagagcagccagccagagcccatctggaagaccccgagcccgcccgcaagacct 1860

```

```

Db 1921 GCAAGGGAGAGACCAAGCCAGAGCCCATCGAGAACCCCAAGCCCCCGCCGAGAGACT 1980
      |||
OY 1861 gcaagggagagcagccagagcccatctggaagaccccgagcccgcaagacct 1920
      |||
Db 1981 GCAAGGGAGAGACCAAGCCAGAGCCCATCGAGAACCCCAAGCCCCCGGAGAGACT 2040
      |||
OY 1921 acaagggatgagccagccagagcccgagggagagcagcagagtctgagagacagaggtgag 1980
      |||
Db 2041 ACAAGGGATGAGCCCAAGAGCGGGGAGAGCCAGACAGTTGCAAGACGAGAGTGGAG 2100
      |||
OY 1981 tctctgcaagcttggaaagccttaagaaagagtgcccgctgcgtctgtgtcctc 2040
      |||
Db 2101 TCTTCTGCAAGTGTGGAGAGCCTTAAGAAAGAGTGCCTGCGGGCTGTGCTCTCC 2160
      |||
OY 2041 tctcctgtcgaagggctggggcctcagagctgtctgtcgaggtccctcaagctctgtc 2100
      |||
Db 2161 TGTCTGTGCAAGGGGCTGGGGCTCGGAGTGTGCTGCGGGCTGCCCTCAGAGCTGTGCT 2220
      |||
OY 2101 tctgtagccgtgaccatctgacccacagtgctgtgctcctgttggggccactagaccca 2160
      |||
Db 2221 TGTGACCCGTGACCCATGACCCACAGTGTGCTGCTCTGTGGGCGCATATAGACCA 2280
      |||
OY 2161 ccagaagccgagagccctcagaggaagcccaagccctgcagaaagcctcctgcgtgctg 2220
      |||
Db 2281 CCAGAAGCCGGAGAGCCCTCAGGAGAACCCCAAGGCTGTGAGAAAGCTCTGAGCTGAGCTG 2340
      |||
OY 2221 tctctcccaacagctctcctcgtcgccctgtcttgaattgaaccttctgagt 2280
      |||
Db 2341 TGTCTTCCCAACCAAGCTTCCCTGCGCCCTGTGTTGTAATTGACCTTCTGTGAGT 2400
      |||
OY 2281 ggggggcgg 2289
      |||
Db 2401 GGGGGGCGG 2409
      |||

```

```

RESULT 4
AF109134 2423 bp mRNA PRI 10-JAN-1999
LOCUS AF109134 Homo sapiens 7-60 mRNA, complete cds.
DEFINITION AF109134.1 GI:4139227
ACCESSION AF109134.1
VERSION AF109134.1
KEYWORDS human.
SOURCE human.
ORGNAMSM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Liu,J., Brewton,R.G., Takanosu,M., Wood,B.M. and Mayne,R.
TITLE Cloning of 7-60: A human gene from chromosome 20q13.3 which encodes
a novel intracellular protein motif repeated seven times
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2423)
AUTHORS Liu,J., Brewton,R.G., Takanosu,M., Wood,B.M. and Mayne,R.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1998) Cell Biology, University of Alabama at
Birmingham, 1670 University Blvd. VHS02, Room 605, Birmingham, AL
35294, USA
FEATURES
source
1..2423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/note="approximately 3kb upstream of COL9A3"
206..2062
/codon_start=1
/product="7-60"
/protein_id="AA03737.1"
/db_xref="GI:4139228"
/translation="WTSRKMRAITRDMCRVHRNTPDLVERDCNGDTPNLSFYRNEIRF
LPNGCFIEDIQQNTDNYDLIEDHASYIOWLPFIREGVWMAKPLTLREVDFKSSQ
EIQERLRAVAHAHGLRDPAGGPRRHGHPSTELPEALPELRSHNDRITRILKS

```


Db	1941	CATGGAGACCCACGAGCCCGGACAGACCTTCAAGGGATAGCCAGCCAGAGGGG	2000
Qy	1946	gsgagcagcagagctctgcagagcagcagagctgagctctctgcacagctctggaaacctt	2005
Db	2001	GGGAGGACGACGAGATTTCAGGAGCGAGAGGTGGAGTCTTCTGCCTCAAGTCTGGAGGCTT	2060
Qy	2006	aaggaagaagatgcacgctcgagctcttgctctctctctctctctctctgcagggctcgggct	2055
Db	2061	AAGGAAAGGAGTGGCCCTCGGCGCTCTTGGTCTCTCTCTCTCTCTCTCTGACGAGGGCTGGGGCC	2120
Qy	2066	tccgagactctgtgaggctctccctccctaaagcctctctctctcgtgagccgctgaacctgaaccac	2125
Db	2121	TCCGAGACTCTCTGGGGCTCCCTCCACAGGCTCTCTCTCTCTCTCTCTGAGCCCGGACCCATGACCCAC	2160
Qy	2126	agtctgtgcctcctctgttgaggcactatagcagccacagaaacgacgagcctctcaagga	2185
Db	2181	AGTCTGTGCTCTCTGTGGGGCCACTATAGCAGCCACCAACACCGGAGGCGCTCTAAGGA	2240
Qy	2186	agcccaagcctctgacagaagcctctctctgtgctgtctctcccaaccaagctctccct	2245
Db	2241	AGCCCAAGGCGCTGACAGAGCCTCTCTGGCGCTGCTGTGCTTCCCAACCAAGCTCTCCCT	2300
Qy	2246	gcgcacctctctctgttaatatgacacctctcgagtgtagggggg	2289
Db	2301	GGCGCCCTGTCTTGTAAATGACCCCTCTGAGAGTGGGGGGGG 2344	

RESULT	5
LOCUS	HS885L7
DEFINITION	HS885L7 160241 bp DNA PRI 09-JUL-2000
ACCESSION	Human DNA sequence from clone RP5-885L7 on chromosome 20q13.2-13.33, complete sequence.
VERSION	AL035669
KEYWORDS	AL035669.43 GI:8979786
SOURCE	HTG.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Enayyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 160241)
JOURNAL	Smith.M.
COMMENT	Direct Submission Submitted (09-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk requests: clonerequests@sanger.ac.uk On Jul 8, 2000 this sequence version replaced gi:8919619. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information from the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/Ref/Chr20 RP5-885L7 is from the library RPci-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://Dacpac.med.buffalo.edu/VECTOR/pcrpac2 This sequence is the entire insert of clone RP5-885L7 The true left end of clone RP4-563E14 is at 140192 in this sequence. The true

```

FEATURES                                right end of clone RP11-93B14 is at 44512 in this sequence.
source                                  Location/Qualifiers
1..160241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.2-13.33"
/clone="RP5-885L7"
/clone_lib="RPC1-5"
2..64
/note="LIM49 repeat: matches 6210..6254 of consensus"
65..353
/note="Alu10 repeat: matches 1..287 of consensus"
354..397
/note="LIM49 repeat: matches 6254..6302 of consensus"
4198..5157
/note="12 copies 80 mer 61% conserved"
/note="LIM49 repeat: matches 6254..6302 of consensus"
/note="Random repeat: region contains two forced joins and
single clone. Assembly consistent with restriction
digest."
4229..5114
/note="443 copies 2 mer CC 59% conserved"
4231..5166
/note="234 copies 4 mer tccc 60% conserved"
4240..5151
/note="19 copies 48 mer 61% conserved"
4248..5159
/note="38 copies 24 mer 61% conserved"
4264..5163
/note="15 copies 60 mer 61% conserved"
4286..5152
/note="17 copies 51 mer 59% conserved"
4308..5171
/note="27 copies 32 mer 61% conserved"
4317..5156
/note="21 copies 40 mer 62% conserved"
4344..5142
/note="17 copies 47 mer 59% conserved"
4348..5145
/note="42 copies 19 mer 60% conserved"
4431..5144
/note="21 copies 34 mer 60% conserved"
4456..5148
/note="21 copies 33 mer 60% conserved"
4486..5151
/note="9 copies 74 mer 69% conserved"
4530..5087
/note="18 copies 31 mer 71% conserved"
4533..4822
/note="5 copies 58 mer 77% conserved"
4720..5137
/note="11 copies 38 mer 61% conserved"
4833..5064
/note="4 copies 58 mer 79% conserved"
5051..5164
/note="19 copies 6 mer tctctc 63% conserved"
5934..6864
/note="LIM4 repeat: matches 4509..5488 of consensus"
6974..7049
/note="4 copies 19 mer 80% conserved"
7224..7451
/note="6 copies 38 mer 98% conserved"
7892..8074
/note="MER8A repeat: matches 68..210 of consensus"
8674..8966
/note="Alu1 repeat: matches 5..296 of consensus"
10876..10962
/note="MIR repeat: matches 86..162 of consensus"
11974..12025
/note="13 copies 4 mer ctca 73% conserved"
14394..14478
/note="5 copies 17 mer 77% conserved"

```

repeat_region	15824. .15891	/note="17 copies 4 mer tccc 67% conserved"
repeat_region	16266. .16577	/note="Alu δ repeat: matches 1. .313 of consensus"
repeat_region	16578. .16866	
repeat_region	16581. .16844	/note="17 copies 17 mer 65% conserved"
repeat_region	16582. .16891	/note="11 copies 24 mer 68% conserved"
repeat_region	16583. .16879	/note="10 copies 31 mer 66% conserved"
repeat_region	16585. .16840	/note="9 copies 33 mer 66% conserved"
repeat_region	16586. .16875	/note="16 copies 16 mer 66% conserved"
repeat_region	16590. .16874	/note="5 copies 58 mer 72% conserved"
repeat_region	16593. .16872	/note="15 copies 19 mer 64% conserved"
repeat_region	16598. .16837	/note="7 copies 40 mer 67% conserved"
repeat_region	16603. .16842	/note="5 copies 48 mer 73% conserved"
repeat_region	16603. .16842	/note="3 copies 80 mer 77% conserved"
repeat_region	16638. .16901	/note="44 copies 6 mer aaagaa 62% conserved"
repeat_region	16664. .16885	
repeat_region	16670. .16904	/note="3 copies 74 mer 78% conserved"
repeat_region	16704. .16899	/note="5 copies 47 mer 77% conserved"
repeat_region	16848. .16899	/note="13 copies 4 mer gaag 100% conserved"
repeat_region	17287. .17411	
repeat_region	17447. .17486	/note="NERF repeat: matches 35. .165 of consensus"
repeat_region	18303. .18502	/note="10 copies 4 mer ccat 92% conserved"
repeat_region	18307. .18498	/note="5 copies 40 mer 66% conserved"
repeat_region	18317. .18496	/note="4 copies 48 mer 67% conserved"
repeat_region	18319. .18486	/note="3 copies 60 mer 74% conserved"
repeat_region	18320. .18495	/note="7 copies 24 mer 66% conserved"
repeat_region	18321. .18496	/note="44 copies 4 mer acac 64% conserved"
repeat_region	18325. .18486	/note="11 copies 16 mer 64% conserved"
repeat_region	18327. .18486	/note="9 copies 18 mer 66% conserved"
repeat_region	18497. .18546	/note="8 copies 20 mer 66% conserved"
repeat_region	19329. .19433	/note="25 copies 2 mer ac 72% conserved"
repeat_region	19455. .19590	/note="TTR16C repeat: matches 286. .387 of consensus"
repeat_region	19457. .19604	/note="4 copies 34 mer 88% conserved"
repeat_region	19546. .19625	/note="2 copies 74 mer 81% conserved"
repeat_region	19839. .19954	/note="4 copies 20 mer 76% conserved"
repeat_region	20020. .20103	/note="2 copies 58 mer 87% conserved"
repeat_region	20704. .20805	/note="MIR repeat: matches 51. .137 of consensus"
repeat_region	20710. .20811	/note="3 copies 34 mer 90% conserved"
repeat_region	20722. .20806	/note="2 copies 51 mer 90% conserved"
repeat_region	21349. .21410	/note="5 copies 17 mer 89% conserved"

[illegible]

Db	62232	CCITCCCGGTCCTCCCTCGGCCATCGGGGAGACCCCAAGCTGGACACAGTGAACAGGGGT	62232
Qy	1533	tgagagagacacagaaggtcgaaacggggcccaagaaggtacccttggagccatcgga	1592
Db	62292	TGAGGAGGACACAGAAGGTCGAACGGGGCCCAAGAAAGGTACCCCTGGGAGCCCATCGGA	62351
Qy	1593	gaacccagggccca-----	1606
Db	62352	GACCCAGGCCCCACAGCAGACTTCAGGGGACGAGCCAGGCCATCGGA	62411
Qy	1607	-----gcacagcagacctcgagggaacgagccagccgaagccatcgga	1652
Db	62412	GACCCAGGCCCCCGCCCGCCAGAGGACCTTCAGGGGACGAGCCAGGCCAGAGCCCATCGGA	62471
Qy	1653	gaacccagggcccgcccgcccgacgacactcgagggaacgagccagccgaagccatcgga	1712
Db	62472	GACCCAGGCCCCCGCCCGGCGAGACTTCAGGGGACGAGCCAGGCCAGAGCCCATCGGA	62531
Qy	1713	gaacccagggcccgaccccgacgacactcaagagatgagccagccgagagcccatcgga	1772
Db	62532	GACCCAGGCCCCACCCCGCGAGACTTACAAGGGGTGAGCCAGCCAGGAGCCCATCGGA	62591
Qy	1773	gaacccagggcccgcccgcccgacgacactcgagggaacgagccagccgaagcccatcgga	1832
Db	62592	GACCCAGGCCCCCGCCCGCGAGACTTCAGGGGACGAGCCAGGCCAGAGCCCATCGGA	62651
Qy	1833	gaacccagggcccgcccgcccgacgacactcgagggaacgagccagccgaagcccatcgga	1892
Db	62652	GACCCAGGCCCCCGCCCGCGAGACTTCAGGGGACGAGCCAGGCCAGAGCCCATCGGA	62711
Qy	1893	gaacccagggcccgaccccgacgacactcaagagatgagccagccgaagcgagggaagc	1952
Db	62712	GACCCAGGCCCCACCCCGCGAGACTTACAAGGGGTGAGCCAGGCCAGGCGGGGAGGC	62771

JOURNAL REFERENCE AUTHORS TITLE JOURNAL SOURCE	Unpublished 2 (bases 1 to 6670) Takanosu, M., Liu, J. and Wayne, R. Direct Submission Submitted (10-DEC-1998) Cell Biology, University of Alabama at Birmingham, VH 302, Room 605, Birmingham, Alabama 35294, USA
FEATURES	Location/Qualifiers 1..6670
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="20" /map="20q13.3" join(<206, 268, 906, .984, 2235, .2313, 3147, .3213, 4125, .4274, 4891, .>6309) 20" /product="7-60" join(206, .268, 906, .984, 2235, .2313, 3147, .3213, 4125, .4274, 4891, .6309) /codon_start=1 /product="7-60" /protein_id="AADO3745.1" /db_xref="GI:4139272" /translation="MTGSRNWRATRDRCRYRHNYPDLVERDCNGDTPNLSFYENLIRE LPNCFITLDMWTDYDLLEDNHSYIOWLFPLRPGVNMHAKPLTLEVEYFKSSQ EIORIVRAAANAGLIRDPAGRRHGGSPSLDEALPEYLRSHNNLRITRLKS LGEELGHEHPAPLVRPELETSYLRRLPCYRQALDYEMWAGCCRRORQIVAAHEH FRPECKTAVGPQDLRLRRFKSSLPPLPEBSGRKYEEEGSPEDPDHEASTGCRICGPHS KGGGRVDGEPQPRSVEPQDAGPLERSQGEGAGGEGDREPLSPKSKRRKLELSRE OPLPEPQGSASEVEKIALNLDEGCALISGSLRTGTQEVGGODPGEAVQPCRPILGARV ADKRRKRRYDEGADSAVAASGAQTLIAGSPADSGHPKAGHSNENVEDETEGRTG PKKETPEPSETPEGPSPAGPADGEPAESPSETRGPSPAGPADGEPAESPSETGPSPA GPNDEDEAKTPSETPGSPAGPADGEPAESPSETRGPSPAGPADGEPAESPSETGPSP PAGPADGEPAESPSETGPSPADPTDRDEPAKGAELADDAVESSAKSGKP"
CDS	1367 a 1998 c 2169 g 1136 t

RESULT	6				
LOCUS	AF112980				
DEFINITION	AF112980	6670 bp	DNA		PRI
ACCESSION	AF112980	7-60 gene, complete cds.			10-JAN-1999
VERSION	AF112980.1	GI:4139271			
KEYWORDS	.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 6670)				
	Takanosu, M., Liu, J. and Mayne, R.				
	Genomic Structure of Human Gene 7-60				

	Query Match	67.9%	Score 1555:	DB 11:	Length 6670:
	Best Local Similarity	95.3%:	Pred. No. 8.2e-20:	Mismatches 20:	Gaps 1;
	Matches 1637:	Conservative 0:	Mismatches 20:	Indels 60:	Gaps 1;
QY	633 ccagaactgacatgcgcgcagaccacaanaacctccgatcatcacgcatacctcaagtgcgc				
Db	4875 CCGGATCTCTTCGCAGGCCACCAACACTTCCGATCATACGGATCTCAAGTGCT				
QY	633 gtgtgagcttgagccctcgagcaacttcaccaagcgcacatggtccgcgttcttcctbgagagac				
Db	4935 GGGTAGCTGGGGCCTCGAGCACTTCAGAGCCCGCTGTCCGCTTTCCCGAGGAGAG				
QY	753 gctgtgtgcggcgaggagcctgcgcgggggtgtgcgcgaatgyccttgtactacttaatttcgc				
Db	4995 CCTGGTGGGGGGGAGCGCTCGGGGGGTCCGGCAAGTGCCTGGACACTTCAATTGATTCGC				
QY	813 cgtgcgtgcgcgaacacaaacgcgcgcacagctgtgtactgtccctggaggacattcgcgcc				
Db	5055 CGTCCGCTGCGGACACCAGCCCGCCACACTGTGTCACTTCGCTTGGAGACACTTCCGCCC				
QY	873 ccgctgcgaagtctgtctgttggggccccaaagaacatctgcgtagttcaagcccagctctct				
Db	5115 CCGCTGCAAGTTGCTGCTGGGGGCCCAAGAACAAGCTGGGAGGTTCAAAGCCACTTCTCT				
QY	933 gccccatcgcgtcgaaggctccagaagaagtgtgaggagaagaagcccccgggagcccca				
Db	5175 GCCGCATCCGCTCGAGGGGCTCCAGGAAGTGTGAAGAGGAAGAGCCCCGGGACCCCGA				
QY	993 ccacgaagcccgagcaecceagggctcggaactgtgtggcccgagagatatgaaggttggggcag				
Db	5235 CCACAGGCGCACACCCAGGGGTCCGACTGTGGGCCAGAGCATGTGCAAGGGTGGGGCAG				
QY	1053 agttgacgaaggagcccccaagccagcagagaggttgaagccccagratgcggaacccctggagag				
Db	5295 GGTGGACAGAGGGGCCCAAGCCACAGGAGCTGTGAAGCCCCAGATCTGGGACCCCTTGGAGAG				
QY	1113 gagccaggggatgagagcagaggggccaagggggaagatagcgcgagacccttaagcccca				

|||||
Db 5355 GAGCAGGGGATGAGGAGGGGCGACGGGAGAAATAGCGCGAGCCCTTAAGCCCAA 5414
Qy 1173 agaagagcaagaagagagagctcggagctgaagccggcgagagcagccgcccacagagcaag 1232
Db 5415 AGAGAGCAAGAAGAAGAGAGCTGAGAGCTGAGCGGGAGCGCCGACAGAGCCAGG 5474
Qy 1223 ccccaagatgctctcagaaggttgagagaatgctctgaatttgagaggggtgtgacctag 1292
Db 5475 CCTTCAGAGTGCCTCAGAGGTGAGAAATGCTCTGAATTTGGAGGGGTGTGCCCTAG 5534
Qy 1293 ccagagcagcctcagagcagggagccacaggaagtggcggtcagagacccctgggagagca 1352
Db 5535 CCAAGGAGCCTCAGAGAGCGGAGACCCAGAAAGTGGCGGTGAGACCTCGGGAGAGCAGT 5594
Qy 1353 gcaagcccgccgcccacccctcgggagcagaggttgccgacaagagtgaagagcagagaa 1412
Db 5595 GCAGCCTGCGCCCAACCCCTGGGAGCCAGGTTGGCCAGCAAGGTGAGGAAGCGAGGAA 5654
Qy 1413 ggtgatatgaggtgtgtgggagacagtgtcgtgtggccagttgtgtgccaacctgtgc 1472
Db 5655 GGTGATATAGGTTGCTGGGGACAGTGTGCGGTGSCAGTGTGTGCTGCCAGACCTTGGC 5714
Qy 1473 ccttgccgggtccctcctcggccatcggggcaccacccagagcttgacacagtgagagcgg 1532
Db 5715 CTTGCGCGGTTCCCTGCTGCGCATCGGGGACCCCAAGGCTGGACACAGTGAAGAGGGGT 5774
Qy 1533 tgagagagacacagagagtgcaacggggcccaagaagtgtaacctgtggagcccatcga 1592
Db 5775 TGAGAGAGACACAGAAAGTCAACGGGGCCCAAGAAAGTACCCCTGGAGCCCATCGGA 5834
Qy 1593 gaccacaagcccca-----gcccagaagacctcagggagcagccagccagagcccatcga 1606
Db 5835 GACCCCAAGCCCAAGCCCAAGACCTGCAAGGGAGACAGCCGGCGAGAGCCCATCGGA 5894
Qy 1607 -----gcccagaagacctcagggagcagccagccagagcccatcga 1652
Db 5895 GACCCCAAGCCCAAGCCCAAGACCTGCAAGGGAGACAGCCAGCCAGAGCCCATCGGA 5954
Qy 1653 gaccacaagcccgcccgccggcagagacctcagggagcagccagccagagcccatcga 1712
Db 5955 GACCCCAAGCCCAAGCCCAAGACCTGCAAGGGAGACAGCCAGCCAGAGCCCATCGGA 6014
Qy 1713 gaccacaagcccgcccgccggcagagacctcacaagagatagccagccagagcccatcga 1772
Db 6015 GACCCCAAGCCCAAGCCCGGAGAGACTTACAAAGGATAGGCCAGCCGAGAGCCCATCGGA 6074
Qy 1773 gaccacaagcccgcccgccggcagagacctcagggagcagccagccagagcccatcga 1832
Db 6075 GACCCCAAGCCCAAGCCCGGAGAGACTTACAAAGGATAGGCCAGCCGAGAGCCCATCGGA 6134
Qy 1833 gaccacaagcccgcccgccggcagagacctcagggagcagccagccagagcccatcga 1892
Db 6135 GACCCCAAGCCCAAGCCCGGAGAGACTTACAAAGGATAGGCCAGCCGAGAGCCCATCGGA 6194
Qy 1893 gaccacaagcccgcccgccggcagagacctcacaagagatagccagccagagccggagagc 1952
Db 6195 GACCCCAAGCCCAAGCCCGGAGAGACTTACAAAGGATAGGCCAGCCGAGAGCCCATCGGA 6254
Qy 1953 agcagagttgcaagagcagaggttgagttctctcgaagttcggaaagccttaagaaa 2012
Db 6255 AGCAGAGTTGACAGAGCAGAGGTGAGTCTTCTCCAAAGTCTGGAAAGCCTTAAGAAA 6314
Qy 2013 ggaatgcccgtcggcgtctgtgctcctctgtcctgtcgcaggggtgtgggacctcga 2072
Db 6315 GGAAGTCCGCTGCGGCTTGTCTCTCTCTGCTGCTGCTGCAAGGGGCTCGGGCTTCGGAG 6374
Qy 2073 ctgctgagaggtcctcctaaggtctgtctgtgaccggtgaccacatgagccacagtggtg 2132
Db 6375 CTGCTGCGGCTCCCTTAGGCTGTGCTTCTGAGCCCTGAGCCCATGACCCCAATGCTG 6434
Qy 2133 gctcctctgtgggcaactatagcaacccagaaagccgagggcccttaaggaagcccaa 2192
|||||

Db 6435 GCTCTGTGGGGCCACTATAGACCAACAGACCCGAGGCGCTCAGGAAAGCCCAA 6494
Qy 2193 gctctgcaagaagcctctcgtgctgtgtcttcccccacacatctccctcgtgccc 2252
Db 6495 GGCCTGACGAAGCCTCTCGGCTGCTGTCTTCCACACACAGCTCTCCCTCGGCC 6554
Qy 2253 tctcttgtaattgacccctcgtgagttggggggcg 2289
Db 6555 TGTCTTCTTAATTGACCTTCTGAGTGGGGGCGG 6591

RESULT 7
AF172450
LOCUS AF172450 1676 bp mRNA PRI 19-APR-2000
DEFINITION Homo sapiens opioid growth factor receptor mRNA, complete cds.
ACCESSION AF172450
VERSION AF172450.1 GI:7595300
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1676)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Cloning, sequencing, chromosomal location, and function of cDNAs
encoding an opioid growth factor receptor (OGFR) in humans
Brain Res. 856 (1-2), 75-83 (2000)
JOURNAL 20143540
MEDLINE 2 (bases 1 to 1676)
REFERENCE Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
AUTHORS Direct Submission
TITLE Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,
Pennsylvania State University College of Medicine, 500 University
Drive, Hershey, PA 17033, USA
JOURNAL
FEATURES
source
1. 1676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.3"
/tissue_type="placenta"
34. 1419
/note="[Met-5]-enkephalin binding protein; OGFR;
alternatively spliced"
/codon_start=1
/product="opioid growth factor receptor"
/protein_id="AF64403.1"
/db_xref="GI:7595301"
/translation="MDPPCDSTWEEDADADDEDCEDGSAAGARADADGDEDE
SEPPRARPSFOSRMTGSRNWRATDMCYRHNYDLYRDCNGDTPNLSFYRNR
FLNPGCTIEDIIONWDNIDYLDLSDNSYIOWLEPLLEPGYNIAKPLTLREVEVFRSS
QEI0ERLVRAVELEMLGTYGRLDRGTVGRKONTQKRQNTNMSSHNLRLTRILK
SPELSLEHQAFLVRFLEETLVRLDPRQSLADYEFVAVRCHOROLVHFAE
HFRPRKFTVWGPDKLRFRKPSLPHLESGSRVLEBGSFSDPHASITGRTCGPEH
SKGGGRVDESDPDRSVPDAPGLPSQSGDAEGHGDRPEPLSPRESKKRLLSR
EOPTEBGPQSAEVEKIALNLGALSQSLTGTQVVGQDPBGAASPCPCGKMWLR
SCGGBSBSGAS"
BASE COUNT 340 a 501 c 575 g 260 t
ORIGIN

Query Match 59.0%; Score 1351; DB 11; Length 1676;
Best Local Similarity 98.6%; Pred. No. 1,4e-189;
Matches 1363; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 tagaatcagcggcgcgtgaattctagccgagcattgagaccccgactgcgactcaacc 60
Db 1 TAGAATTCAGGGGCGCTCGATTTCTACCGAGCATGAGACACCCCGACTGGAATCCACC 60
Qy 61 ttggaggaaggaacgaagagatgacgaggaacgacgaggaacgaagacctgcgagagcgag 120
Db 61 TGGAGAGGAGCAGAGAGATGCGAGAGACCGGAGAGACGAGACTCGAGAGCGCGAG 120
|||||

[illegible]

Db	1201	ACCCGCGGAGACACCCGCCACAGAGCCAGCCCTCAGATGCTCTCAGAGGTGAGAG	1260
Qy	1261	atcgcctcgaatttgagagggtgtgtccctcagccagcagcagcctcagcagcggagaccag	1320
Db	1261	ATCGCTGTGAATTGTGAGAGGGGTGTCCCTCAGCCAGGCGACCTCAGAGACGGGAGACCAG	1320
Qy	1321	gaagtcggcgcgtcagagacctggggagcagtcagccctgcgcgaacctctggagac	1360
Db	1321	GAAgTGGCGCGGTcAGAGACCTGggGAGGCGTcGTGCTCTCTCTCAGAGGgGCTGGGgGCTC	1360
Qy	1381	agg 1383	
Db	1381	CGG 1383	
RESULT	8		
LOCUS	AF172449		
DEFINITION	AF172449 1232 bp mRNA PRI 19-APR-2000		
ACCESSION	AF172449		
VERSION	AF172449.1		
KEYWORDS	GI:7595298		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 1232)		
JOURNAL	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.		
MEDLINE	Cloning, sequencing, chromosomal location, and function of cDNAS		
AUTHORS	encoding an opiod growth factor receptor (OGFR) in humans		
JOURNAL	Brain Res. 856 (1-2), 75-83 (2000)		
REFERENCE	20143540		
AUTHORS	2 (bases 1 to 1232)		
TITLE	Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.		
JOURNAL	Direct Submission		
FEATURES	Submitted (26-JUL-1999) Department of Neuroscience and Anatomy,		
SOURCE	Pennsylvania State University College of Medicine, 500 University		
	Drive, Hershey, PA 17033, USA		
	Location/Qualifiers		
	1..1232		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="20"		
	/map="20q13.3"		
	/clone="127"		
	/tissue_type="placenta"		
	34..1212		
	/note="[Met-5]-enkephalin binding protein; OGFR;		
	alternatively spliced"		
	/codon_start=1		
	/product="Opiod growth factor receptor"		
	/protein_id="AAFG4402.1"		
	/db_xref="GI:7595298"		
	/translation="MDPPDQSTWEEDENADPADPDCEDEAGARADACDEDEE		
	SEEPRARAPSSFRNMGSRNNRATRDCKRYINPDLYVRDNGNDPINTSTRNIR		
	FLBPGFTIEDILQNDWNYDLDDNSYIQWLEPLDEPGVNNHAKPLTLREVEFKSS		
	QEOIERLVRAVELMELGYGIRLDRGTGTGRAQNTQKRFQNLNWRSHNRLRTIRLIK		
	SPEGLSTLEHQAQVLVRFLEETLVRELPVRSALDYEFMFAVCRSHORQLVFAWE		
	HFRRCFTVWGPODKLRFKRFSSLPPLHLEGRKVEEGPAGDEPASPSPGSPAG		
	PTNREPKAGAEAEACCLVAHSVHLPALCAPIFVNRPPWMSGGRAGLAFSLMPKASE		
	"		
BASE COUNT	256 a 369 c 403 g 204 t		
ORIGIN			
Query Match	42.7%: Score 977.4; DB 11; Length 1232;		
Best Local Similarity	99.4%: Pred. No. 1.1e-134;		
Matches 981: Conservative 0; Mismatches 6; Indels 0; Gaps 0;			
1 tagaattcagcgcgcgtcgaattcttagccgagcaltgacgaccccgactgcacc 60			

Db 1 TAGAATTCAGCGCGCGCTGAATTCAGCCGAGCATGAGACGCCCGGACTGCGACTCCACC 60
Oy 61 tggagagagagagagagagatgcgagagacgagagagagagagagagagagagagagagag 120
Db 61 TGGAGAGAGAGAGAGAGAGATGCGAGGAGCGGAGAGACGAGAGACTGCGAGGAGCGGAG 120
Oy 121 gccgc 180
Db 121 GCCCGCGCGCGCGAGGAGCGCGAGCCAGGAGGAGAGAGAGAGATCCGAGAGAGCGCGCG 180
Oy 181 gc 240
Db 181 GCGGCGCGCGCGCGAGCTCTCCAGTCCAGAAATGACAGGGGTCCAGAACTGCGGAGCCACG 240
Oy 241 agggagcagtgtagtgc 300
Db 241 AGGGAGCAGTGTAGTATCGGACACATCTCCGAGTCTGGTGAAGACGAGCTGCATGGG 300
Oy 301 gacacgcacaaacctgaattctacagaaatgatcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
Db 301 GACACGCCAAACCTGAGTTTCTACAGAAATGATGATCCGTTCTCTGCCAACGCGTGTTC 360
Oy 361 attgagagacatcttcgaagactggaacaaactatgacctctctgaggacaactctcc 420
Db 361 ATTGAGGACATCTTCAGAACTGAGACGAGACACTATGACCTCTGAGGACAACTCATCTCC 420
Oy 421 tacatccagtgctgttctctctctgagagacagagagatggaactggtcagccacccctc 480
Db 421 TACATCCAGTGTGCTTTCTCTCTGAGAGAACGAGAGTGAATGCGCATGCGCAAGCCCTTC 480
Oy 481 acgcgcagagagagtgatggttttaaaagctccgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ACGCTCAGGAGAGTGAAGGTGTTTAAAGCTCCAGAGATCTCAGAGAGCGGCTTCTCCGG 540
Oy 541 gctcagagactcagtcggtctctcagcgagatccgcgcgcgcgcgcgcgcgcgcgcgcgc 600
Db 541 GCTCAGGAGCTCATCTGAGGCTTCTACGAGGATCCGCGTGAAGACCGAGGAGCGGCGACG 600
Oy 601 gtgggc 660
Db 601 GTGGGCCGAGACAGAACTACCAAGAGCGCTTCCAGAACCTGAATGCGGCGGCGCACAAAC 660
Oy 661 aacctccgcacacacgc 720
Db 661 AACCTCCGATCACACGCATCTCTCAAGTCCGCGTGTGAGCTTGAAGCTTGAAGCTTCCAG 720
Oy 721 gc 780
Db 721 GCGGCACGTGTCGCTTCTCTCTGAGGAGACGTGTGCGGCGGAGCTGCGGCGGAGTGC 780
Oy 781 cggcagagtgccctggaactactcattcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 781 CCGCGAGATGCTCTGAGTACTTCAATGTCGCGCTGCGCTGCGACACAGCGCGCGCAG 840
Oy 841 ctggtgcaccttcgc 900
Db 841 CTGCTGCACTTCCTGCGGAGCATCTCCGCGCCCTGCAAACTTGTCTGGGGGCGCCAA 900
Oy 901 gacaaagctgc 960
Db 901 GACAAAGCTGCGGAGGTTCAGAGCCAGCTCTGCGCATCCGCTGAGAGGCTCCAGGAAG 960
Oy 961 gtggag 987
Db 961 GTGGAGGAG 987

RESULT 9
AF156878 2208 bp mRNA ROD 16-JAN-2000
LOCUS AF156878 Ratius norvegicus oploid growth factor receptor mRNA, complete cds.
DEFINITION AF156878
ACCESSION AF156878
VERSION AF156878.1 GI:6693831

KEYWORDS
SOURCE Norway rat.
ORGANISM Ratius norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Ratius.
REFERENCE 1 (bases 1 to 2208)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Cloning, sequencing, expression and function of a cDNA encoding a receptor for the oploid growth factor, [Met(5)]enkephalin
Brain Res. 849 (1-2), 147-154 (1999)
JOURNAL MEDLINE 20074706
REFERENCE 2 (bases 1 to 2208)
Zagon, I.S., Verderame, M.F., Allen, S.S. and McLaughlin, P.J.
Direct Submission
Submitted (04-JUN-1999) Neuroscience and Anatomy, Penn State University College of Medicine, 500 University Drive, Hershey, PA 17033, USA
FEATURES
source location/Qualifiers
1..2208
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="18 day-old fetus"
151..1893
/note="[met-5]-enkephalin binding protein"
/codon_start=1
/product="oploid growth factor receptor"
/protein_id="AAF25002.1"
/db_xref="GI:6693832"
/translation="MDPPDCSTWEESEEDGEDGQADPTDEDTGDDDDAARARS
LFQSRMTGIRKRAMQDMQRIKRENTPLTQDDCNDGDCNLSFKINELCFQPNGLIED
ILOMKNDYDLLEENHSYIQWLPRLPGVNHAKPLTLEVEAFKSKVREKLVYA
YELMGFYGHLEDRGTGAVCRQNPQPREHNLNHSNMLRTRILKISGELLEY
QAPLVREFLEETLVQHKLPVROGALDYFLFAVCRHORELYFAHEHFKPREFYW
GPRDKLRRFPORITPOPLTGPQADKREGSDQEAQTGRTGCGSGRDLSSGTAE
DPSLNTKRPDGGTLDGNQDDEAKSLSPKSKRKRLKLEGNQBOVPAGADQGVSEVK
IALMLECALSPISQEPREAPPCPVARVANEVRRKRVKVEGAGDGVSNQWQABA
LPPTPSECPRAQKNGNDEPNSQVGAEDSKSOVGPDPNSQVLEDPNSQVGEDPN
SONGPEDPNSQVGEDPNSQVGPDPNSQVGPDPNSQVGPDPNSQVGPDPNSQVGPDPNSQV
ELARIEASAEPRKP"
repeat_region 1549..1764
/note="imperfect"
/rpt_type=direct
/rpt_unit=1549..1575
2031..2036
polya_signal
polya_site 2203
BASE COUNT 522 a 615 c 675 g 396 t
ORIGIN
Query Match 30.5%; Score 698.4; DB 12; Length 2208;
Best Local Similarity 70.0%; Pred. No. 8.4e-94;
Matches 1046; Conservative 0; Mismatches 376; Indels 72; Gaps 5;
Oy 48 ctgcgcacccactgc 107
Db 135 CTCTGCTTGTGTCGATGAG 194
Oy 108 cgaag 167
Db 195 GGAGGATGGCGAGGATGGCGAGCGAGGATGATGACGACCATAGGAGACAGCGGCGAGATGA 254
Oy 168 ggaagagc 227
Db 255 CGCGCAGCGCGGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAA 314
Oy 228 ctgcgcagc 287
Db 315 CTGCGGTCTGTATGACGAGCATGCAAAAGATACCGGACAACTACCGGATTTGACAGATCA 374
Oy 288 agactgcacatggtgggc 347

||||| 375 AACACTGCATGGTGCATATGCAACTGAGCTTCTCAAAAATGATATGCTTCCAGCC 434
||||| 348 caaagcgtgttcaatctgaggaacattctcagaactgagagcaacatgacctctga 407
||||| 435 AATATGGGCTCTCATGAGGACATCTCTCAGAACTGGAACACATATGACTCTCTGGA 494
||||| 408 ggaacatactctctacatcagtggtgttctctctgagagaaacagaggtgaactggga 467
||||| 495 AAGAAATCACTCCTACATCCAGTGGCTGTTCTCTGCGGGAACGAGAGTGAAGTGGCA 554
||||| 468 tgcgaagccctcaccgctcagggaggctcagagtgtttaaagctcccaagagatcccaaga 527
||||| 555 CCGCAAGCCCTCACCCTAAGAGAGGTAGAGATTTAAAGTCCAAAGAGAGTGCAGAGA 614
||||| 528 ggcggtctgtccgggacctacagagctcagctggtggtctctcagggatccggctggaagccg 587
||||| 615 GCGTCTTGTCCGGGCTATGAGCTCATGCTGGGCTTCTATGGGTTCCACCTTGAGAGCCG 674
||||| 588 aggcacgggcaaggtgtggcgagcagagactaccgaagggcttcacaaactgagctg 647
||||| 675 GGGACAGGGTGTCTGATGCGGTGCACAGAACTTCCAGCCGCTTCCACAAATCTGAACAG 734
||||| 648 ggcgaagccacaacaacctccgcacatcacagagctcctcaagctcgctgtgagctgagcct 707
||||| 735 CCAGAGCCACACAACTCCTGATTAACAGCATCTCAAGTCACTGAGAGTGGGCTT 794
||||| 708 cgaagactccagagcgacagctgctcgtctctctctcagagagagcagctgtgcgaggga 767
||||| 795 AAGAACTCACTCAAGGACCCCTGCTCGCTTCTCTGAGAGAGACCTTGTATCAGACAA 854
||||| 768 gctgcgggggtgtggcgagagagtgccctgagactcaactcagctgcgctgctgcggaga 827
||||| 855 ACTGCCACAGTGCAGCAAGTGCCCTGACATCTTCTGTTGCTGCTGCTGCTGCTGCGGGA 914
||||| 828 ccaagcgcccgacagctgtgtcactctgcctcggaggagcacttcggccggcgctgcaagtctg 887
||||| 915 CCAGGCGCGGAGCTGTGTACTTGTCTGGGAGCACTTCAAGCCCTCGCGAGAGATTGCT 974
||||| 888 ctggggggccccaagacaagctgcgagagttcaagcccagctctctgcccactcgcgtga 947
||||| 975 CTGGGGGGCCCCCTGACAACTCGGAGATTCAGGCCCCAGACCATACCCAGCCACTGAC 1034
||||| 948 ggggtccccaagaggtgtgag 1007
||||| 1035 GGGACCAAGGCGAGCGATGAAGATGAGGCTCCAGAGACCCCTCCCAAGAGGCTGGCAC 1094
||||| 1008 ccaaggtcagagactgtgagcgagagatagcaagggctgaggagagagagagagagagag 1064
||||| 1095 CCAGGGTCCGACTGTGATGGAAGGACCTGAGTGGGAGACAGTGAAGTGAAGTGAAGTGA 1154
||||| 1065 gcccccagcag 1124
||||| 1155 TCCCTCACATGCTGAACAAACCCCTCAAGTGGGGGAACCTTGGAGGGAACAGAGGGA 1214
||||| 1125 tgaagcagaggggccaacggagagatagcgccgaagcccttaagccccaagagagagagag 1184
||||| 1215 TGAAG-----CTAAGTCCCTGAGTCCCAAGAGAGCAAGAA 1250
||||| 1185 ggaagagctggaactgagccggcgagagcagccgccaagagcagagccctcagaagtgc 1244
||||| 1251 AAGGAAGTTGAGAGGGGAACAGGAGAGAGTCCAGAGGAGAGAGTCCCAAGGAGGTGT 1310
||||| 1245 ctcaagaggtgagagagagatcgctctgaatttgagaggggtgtgcctcaagccaagggcagcct 1304
||||| 1311 CTCTGAGTAGAAGAAATTTGCCCTTAACCTTGAGAGTGTGCCCTTATCAGC-- 1368
||||| 1305 caggaagggagaccaggaaggtggcggtcaggaacctggggagagagagtgcaagccctgccc 1364
||||| 1369 -----CAGGAGCCCAAGGAGAGGCTGAACCCGCTGTCC 1400
||||| 1365 ccaacccttgagagccaaggttgccgacaaagtgaagaaagcgagagagagtgatagagg 1424

Db 1401 T-----GTGGCCAGGTTGGCTATGAGGTAAAGAAAGCGAGAGAGTGGAGAAAG 1451
Qy 1425 tgcctggagacagctcgtcggtgagccagtgtgtggtgagccagacctgacctgaggtc 1484
Db 1452 GGCTAGAGGTGATGAGTAGT-----CAGTAACACTCAAAATGCAAGGCAAGTGGCCGCTGCC 1505
Qy 1485 cccctggccatcgaggagcccaagggctgagacagagtgagagagagagagagagagag 1538
Db 1506 TCTTACCCCTTCAAGTGTCTCTGAGGCCCAAAAGATGGAGTGGCCAGAGGA 1559

RESULT 10
HHU9228/c
LOCUS
DEFINITION
Human herpesvirus 6 serotype B putative major immediate-early
genes.
ACCESSION
VERSION
U92288
U92288.1 GI:2769711
KEYWORDS
SOURCE
ORGANISM
Human herpesvirus 6.
Human herpesvirus 6
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Betaherpesvirinae; Roseolovirus.

REFERENCE
AUTHORS
TITLE
Kosuge, H., Isegawa, Y. and Yamanishi, K.
1 (bases 1 to 30752)
Nucleotide sequence analysis of a 30-kilobase-pair region of human
herpesvirus-6B (HHV-6B) genome and strain-specific variations in
major immediate-early genes
Virus Res. 52 (1), 1-14 (1997)
98115230

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
Kosuge, H., Isegawa, Y. and Yamanishi, K.
2 (bases 1 to 30752)
Submitted (07-MAR-1997) Virology, Research Institute for Microbial
Diseases, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565, Japan
Location/Qualifiers
1. .30752
/organism="Human herpesvirus 6"
/serotype="B"
/isolate="HST"
/db_xref="taxon:10368"
<1. .1281
/gene="H74"
<1. .1281
/gene="H74"
/function="replication, helicase-primase complex"
/note="part of helicase-primase complex, similar to HHV6
U74, HCMV UL102, HVS gene 41, EBV BBLF3, EBV1 54, VZV 52,
HSV1 UL8"
/codon_start=1
/product="H74"
/protein_id="AAC40324.1"
/db_xref="GI:2769712"

/translation="LQSVFAFLHEKIRGVYKQVLYQICEYIGDPLMFEGNERSVFTG
YPLMLTFSVDSLERRVDPDTYICREILISFGLAIPILGRNHAIPVIRELSVEMPGSE
TSLDRFPNSOYVSESLCFTQTEGDETHLFFPSDSMDVTLPLDCLRLILSTVPAFL
PCPENATREILIKEMSRLOHRSYALFVEMLDFAVFAFRACITLMRMRLVLD
LHVYTLTCDGRKDSVYVPLDTATNENMEKITEIKOPARQCHESISGCVFAREFL
SSGLSQRKEMWYATASITACIRANKTGLCFRSTIVITQDMACVFATIRINYSVI
EEEPDNYCEYHEGLDCSRVYAVMSDDQLVRQAGIALTDKINSYIILVTLRL
RRWSEKFEDEVQTEFIRMENRMLEYEHIHLHLN"
complement(1316. .2065)
/gene="H75"
complement(1316. .2065)
/gene="H75"
/note="similar to HHV6A U75, HCMV UL103, HVS gene 42, EBV
BBLF2, EBV1 55, VZV 53, HSV1 UL7"
/codon_start=1
/product="H75"
/protein_id="AAC40325.1"
/db_xref="GI:2769713"
/translation="MKSILKRLKLETSVDVENTLHVITILVIAIKDCVSLARSLVNI
TVGDGIMFHLDEDTMINGLEYKTIKEKELFQGFQIILIDSEDTLOELRLNPFQK

CDS

gene

gene
CDS
/gene="H76"
/function="possible viron protein"
/note="similar to HHV6A U76, HCMV UL104, HVS gene 43, EBV BBRF1, EBV1 56, VZV 54, HSV1 UL6"
/codon_start=1
/product="H76"
/protein_id="AAC40326.1"
/db_xref="GI:2769714"
/translation="MHASANSGLNSVSGMMWRROSSGRRSRKSLSDNEATLSTINS ILGAPDMSKSLSTLPNNSEIDMTYSEQIMTITEMHGKNEFKGOTINALRDS ALVKQIAIAGVAQALLNSYSIOIHDENKRVHRSPEFNKRLSEFQDYFSVMAHAKQV ILGDISINIFNLIQIDNSHYRVDMICTGIVPEMRITTPAPMLNLQOVSRLI HDIVHKONIVPILGLSSVLIIPDFHNKIPRDNSEQISCFKNNKAIAPFTYSTPY VRRNRLMTTPLAHLSPDLKNSLRROKMOCLNTPETIKVLTAKDVTAKKIMDM IEKEKNSDAKSLIKFLILSDSKSIGRDSVSGFTIOETIPSTIIDONKMLNGOF BRKSAIDIGERVDRLFKQITIKCMEBOIQOMDEITETKTNOFPERKIKHLSLE TNNCCDRINPDLDHDLNLSRALNIVQRLFTSVSIDDTSVANSFSSQYIPYQY ADKRIDQJMEMEYMTFRLKKNVNOQGESITYNSITELIIVPELRRLNLYMES IPEEFLFLSIGEILAIYESSKIKHYRLVYVRELNOISEVYNLTQTHPENNEPIFDS NIFSPNEEIELEKIKRIRNLRIOHLTRPNPKGDQD"
3757..6231
/gene="H77"
3757..6231
/gene="H77"
/function="helicase, helicase-primase complex"
/note="similar to HHV6A U77, HCMV UL105, HVS gene 44, EBV BBLF4, EBV1 57, VZV 55, HSV1 UL5"
/codon_start=1
/product="H77"
/protein_id="AAC40327.1"
/db_xref="GI:2769715"
/translation="MSISLFGGRYDNKFLNMSAPRIELIVDKVASLSERRLGR L PEDMRHIMDPETERNSEPADALCIGIDENQPLPRLPKALLYGTGACGTNIGT LAANDCTVATATSTIAQNLSTVLRKSAQKTLFKTFGNSHYVSSERSOYIAND EBSIOIQKODSITWNVLSIDIAERALGVAQKTELPDLCESSIVIDEAVILLRI LHTVFFWFWYNAKLTPLYENGIVPCVQSPQSNALVSPNLPLOKCKVKGID VLSALICDVLSTKCEVDNMMWIIFVNKKCADHAFGDFLKHIEFGILPKPELIEVDQ EYKPSAYIRNPNNEIETRLFLSHNEKNYERLSEHOYEVTNRNMLFEPVYELFNK TFEYKSEIGNSLIEPEWKSNIHRLNMYSOQADODLSKYOLEIYELIOWSVEETL ITCHLKHTRNSSIGVTSKIACTVCSGTETKEFVLLQSDLEIKTSCBQITHAYSFL SGLMFGMSTSCSEFTTPEVLMETKNIKMPSEIELESMSMSRDVOVETDERYDF GLVDDGLSDMDLEIDPCGDPEFTFRYSKILPNSLSFEIISILYTFDFIDIFSRAIL OMHTKGFEGKTLTVYNNRNVSRRKOCGEIYSHLKSFGMLTYAIPANNYTEGYNDN VVHLGTDKQIPQILKKGLPLVVKDEMGFISVLNNYSKFDVYVNGOSFHTCTVDY ATSKVSMITKSGSLSTOKVADRGSDPKNLKLSIYVGMARVTDPNMLNANVPLR LNVENDNFIAPHTVAKLNENTMLIF"
6495..6722
/gene="CB5RH"
6495..6722
/gene="CB5RH"
/codon_start=1
/product="CB5RH"
/protein_id="AAC40342.1"
/db_xref="GI:2769730"
/translation="MTSAKTNBSLPRSTRMYSGNMLKNAQFSLSSRELVNSFYDVS GGGLDGLVLSINSSFIPTTEYILGIVRIR"
8179..8406
/gene="CB6RH"
8179..8406
/gene="CB6RH"
/codon_start=1
/product="CB6RH"
/protein_id="AAC40343.1"
/db_xref="GI:2769731"
/translation="MLHNVSKICHSICIRVCIKLHVICSSRPSIRCAFAVYETYSILPN TSDGRQAFLLYAYGKVHALVDAIRKGRFG"
8829..9866
gene

CDS
/gene="H79"
8829..9866
/gene="H79"
/function="possible replication protein"
/note="similar to HHV6A U79, HCMV UL112, spliced in vitro replication protein"
/codon_start=1
/product="H79"
/protein_id="AAC40328.1"
/db_xref="GI:2769716"
/translation="MYAEKRGYGSFQNYQAYEQIISQSLHLKRFEPDNGCFIEFLAD SGTCETFEKGMISMYWTSETDSMGLTVDIGMEGKRTYRARGALLCSKSTISIQ NTEGERILTVSHENGKLOITFVTAKVSSPELRLNGDLKEMKEFEKCALRKH DDEHRRSGKOKERKVDIDDKKDEKLODEKRNEDKRPDKKDEFGKCYAVV VNRFTIGSRSMFTPIVNSICFYVISPPEPRKOKSHETKRNLEDSHDGIAPSTTF VNGAVGALSFCVSTIDNHEDQHDRLRYVAQGVGLGSPRPSLLELCVSRADLF LIGDEPRRSIQ"
9838..10449
/gene="H80"
9838..10449
/gene="H80"
/function="possible replication protein"
/note="start-codon 99g; similar to HHV6A U80, HCMV UL113, spliced in vitro replication protein"
/codon_start=1
/transl_except="(pos:9838..9840,aa:Met)
/product="H80"
/protein_id="AAC40329.2"
/db_xref="GI:7687902"
/translation="MTNREGLSNEDNYGNFOLNLSLEQRLARLVASSGEVERSLK LERLDYVDNLIKKNVLECADVTVSKSKTKRHEKOKIOIYFSDVRSVPCETKPI DMRVETETTONARVRVROTRRTVSGTSGAIGQORVITSGQRRARGRGRVPRRN SNLNLRQNSAIYVIDSSETEENENASFNEDLATITETL"
complement(10493..11260)
/gene="H81"
complement(10493..11260)
/gene="H81"
/note="similar to HHV6A U81, HCMV UL114, HVS gene 46, EBV BKR3, EBV1 61, VZV 59, HSV1 UL2"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="AAC40330.1"
/db_xref="GI:2769718"
/translation="MALLQWMLDHYODEKRYENSLDQHSIFGINDWLSFLOLSK LEITHKAVYLVNDRAHLTVHPSDSNVHAASTLCKTQVYVILGHDPDGGCG LAFGTVCBSIPESLKNTFKELERSIPNESPDMGCLNSWCREGVLLNSLFTVHGL PMSHAFGMQWLSYKILSRLEQONSIVFLMGHARLSTIDAOKHLYLESAPSP KYKARMPFICNHFVRNLTFLTEHGKDPIMNNILNE"
complement(11336..12088)
/gene="H82"
complement(11336..12088)
/gene="H82"
/function="gH-gL complex, infection, cell fusion"
/note="gH; similar to HHV6A U82, HCMV UL115, HVS gene 47, EBV BKR2, EBV1 62, VZV 60, HSV1 UL1"
/codon_start=1
/product="glycoprotein L"
/protein_id="AAC40331.1"
/db_xref="GI:2769719"
/translation="MELLFVMSLLILTFSSKAMPLEHDNHSFYFEKLDLDCIAVINCTR SEVPLLEPIYQAPVYNEDVMSILKPKTKKPSRIWNTPELILQDNEQVNI TLFALIGPESRDWNLNFRNFGQCSVSGITTCISDNCKRYLDERITYVANNFYDNI AGLEFNISLQNDSTFSYINIGFLYLENPATGILKTIIRFPNSLTLEDTILNCKLYFHK TGEVFDLQKQMEAYNSKLPFRSSRPTILIRMT"
12057..12506
/gene="H83"
12057..12506
/gene="H83"
12057..12506
gene
CDS
Query Match 4.6%; Score 106.4; DB 72; Length 30752;
Best Local Similarity 55.4%; Pred. No. 2.4e-07;
Matches 206; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
QY 1572 taacctgtgagaccatcgagaccagcccccagaccagcctgcagggaaga 1631

Db	15883	TCCAGCAGAGCCTTCACGACAGAGCCTCCACGACAGCCTCCACGAGGGCTCCACGAGAGGCC	16824
QY	1532	gccagccgagagcccatctgagagacccccagcccccgccccgagaggaactctgaggygagca	1691
Db	16823	TCCAGCAGGGGCTTCACGACAGGGCTTCACACAGGCTTCACGAGGGCTTCGGTAGGGCT	16764
QY	1692	gccagccgagagaccatctgagagaccccaagcccccaagccccgagaggaactacaagagatga	1751
Db	16763	TCCAGCAAACTTCACGCGGAGTCTTACGAAAGCTTCACGAGGGCTTCAGAAAGGCC	16704
QY	1752	gccagccgagagaccatctgagagaccccaagcccccgccccgagaggaactgaggygagca	1811
Db	16703	TCCAGCAGGGCTTCACGACAGGCTTCACAGAGCTTCACGAGAGGCTTCACAGAGGGCC	16644
QY	1812	gccagccgagagaccatctgagagaccccaagcccccgccccgagaggaactgtagagagca	1871
Db	16643	TCCAGCAGACTTCACGACAGGCTTCACACAGGGCTTCACGAGAGGCTTCACGACAGACC	16584
QY	1872	gccagccgagagaccatctgagagaccccaagcccccaagccccgagaggaactacaagagatga	1931
Db	16583	TCCAGCAGAGCCTTCACGACAGGCTTCACAGAGGGCTTCACGAGGGCTTCACGAGAGCC	16524
QY	1932	gccagccgaagc	1943
Db	16523	TCCAGCAGAGCC	16512

[illegible]

RESULT	11
LOCUS	AB021506/c
DEFINITION	AB021506 161573 bp DNA VRL 14-APR-2000 Human herpesvirus-6 (HHV-6), variant:B, strain:HST, viron genomic DNA, complete sequence.
ACCESSION	AB021506
VERSION	AB021506.1 GI:4995977
KEYWORDS	.
SOURCE	Human herpesvirus 6 (strain:HST, pop_variant:B) DNA. Human herpesvirus 6 (strain:HST, pop_variant:B) DNA. Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae; Roseolovirus.
REFERENCE	1 (sites)
AUTHORS	Isegawa,Y., Mukai,T., Nakano,K., Kagawa,M., Chen,J., Mori,Y., Sugawara,T., Kawashishi,K., Seshihara,J., Hata,A., Zou,P., Kosuge,H. and Yamanishi,K.
TITLE	Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B
JOURNAL	J. Virol. 73 (10), 8053-8063 (1999)
MEDLINE	99412319
REFERENCE	2 (bases 1 to 161573)
AUTHORS	Isegawa,Y.
TITLE	Direct Submission
JOURNAL	Submitted (13-DEC-1998) to the DDBJ/EMBL/GenBank databases. Yuj1 Isegawa, Osaka University Medical School, Department of Microbiology; Yamada-oka 2-2, Suita, Osaka 565-0871, Japan (E-mail:i-segawa@micro.med.osaka-u.ac.jp, Tel:81-6-879-3323, Fax:81-6-879-3329)
FEATURES	location/Qualifiers
source	1..161573 /organism="Human herpesvirus 6" /virus="" /strain="HST" /db_xref="taxon:10368" /pop_variant="B" complement(18..365) /gene="LTI" complement(18..365) /gene="LTI" /note="52.5% identical to LTI gene of strain UI102 of HHV-6" /codon_start=1 /protein_id="BA78211.1" /db_xref="GI:4995978" /translation="MVRARVRAMVSAMRVAMRVVRVMRAMRVAMRVAMRVRRVVR"
gene	
CDS	

[illegible]

```

gene
6512..7150
/feature="transforming activity"
/feature="42.28 identical to DR7 gene of strain U1102 of
HHV-6
US22 gene family"
/codon_start=1
/product="transactivator"
/protein_id="BAA78217.1"
/db_xref="GI:4995984"
/translation="MRAARLCIASNVTAEFARRGIQADPVAVEEDLEIDPRRMCGTNN
RLFLFIAAAGEHDLFTVGGICQTHAGVSCLEETDPTAVPGVMTLTVAAQYV
RLMPADRLRIHLIYAGHGLSPWICAVLSREHQTSPISGGVTLGAVPPGRR
EVE7AMVLYTWMDRCYRSRPTIARSAVWRKSPGCGMAAGHETLDFGCG"
complement(7236..7706)
/feature="DRH2"
complement(7236..7706)
/feature="DRH2"
/feature="ORF unique to HST strain"
/codon_start=1
/protein_id="BAA78218.1"
/db_xref="GI:4995985"
/translation="MLAPRTPGVYIKECGLASSPDLCDSPGVGIRGVCYSANGRG
RCLIDRKVFFELCGVSNTEPKHIVCCVSPILFLGLAGDTVRYRTHTHRR
RRIDISTAVLELFLRLRTRPHFFVFLPEILVMTFLSCMLPIASRLARGV"
7928..8662
/feature="DR8"
7928..8662
/feature="DR8"
/feature="13.68 identical to DR8 gene of strain U1102 of
HHV-6"
/codon_start=1
/protein_id="BAA78219.1"
/db_xref="GI:4995986"
/translation="MROVTRDDKGRLLAWOLGKKRQCKILTKETSLPLTPPSG
TIPPLRNHFCRFSRHRVYQOTRRHHTSAMVAGGAGSIRGADVSMNGIARVY
BSRRLATILARRQKGLASITLTLTLTLTLTLTLTLTLTLTLTLTLTLTLTL
LTLTLTLTLTLTLHPPTBARAPLMEACFSPKRAPLEAREKTSPPERARVACRR
CAPRPSPLPLPPKRGKRGYGGGVRV"
complement(8292..8807)
/feature="LJ1"
complement(8292..8807)
/feature="LJ1"
complement(8292..8807)
/feature="LJ1"
/feature="16.68 identical to L1 gene of strain U1102 of
HHV-6"
/codon_start=1
/protein_id="BAA78220.1"
/db_xref="GI:4995987"
/translation="WALAIVLSFLRLLVCPBSALPSASLSLSRAPSFRRHAPADGC
CFAPRLAHIPHTLARSIGGGGGIPAGRGDTGRGLGVREGRFHAALAARVAV
VNTAPPIEARRALADGIGLIGLIGLIGLIGLIGLIGLIGLIGLIGLIGLIGL
LGLIGLIGLIGL"
8929..9384
/feature="U1"
8929..9384
/feature="U1"
/feature="25.68 identical to U1 gene of strain U1102 of
HHV-6"
/codon_start=1
/protein_id="BAA78221.1"
/db_xref="GI:4995988"
/translation="MRAAPRAIDISDPRIYSWQIIRTRTDPTPRHRAOVHSE
SIHTEKIRTKYQDENTQKTYHKKIHTQKIHKKIHTQKTYHKKIHTQKTYH
HKKTYHTKKYHTKKYHTKKIHTKKIHTKKIHTKKIHTKKIHTKKIHTKKIHT
complement(9467..10768)
/feature="U2"
complement(9467..10768)
/feature="U2"
complement(9467..10768)
/feature="92% identical to U2 of strain U1102 of HHV-6
US22 gene family"
/codon_start=1
/protein_id="BAA78222.1"
/db_xref="GI:4995989

```

[illegible]

AUTHORS Lindquester,G.J., Inoue,N., Allen,R.D., Castellini,J.W., Stamey,F.R., Dambaugh,T.R., O'Brian,J.J., Danovich,R.M., Frankel,N. and Pellett,P.E.
TITLE Restriction endonuclease mapping and molecular cloning of the human herpesvirus 6 variant B strain Z29 genome
JOURNAL Arch. Virol. 141 (2), 367-379 (1996)
MEDLINE 96195263
AUTHORS 4 (bases 64106 to 84963)
TITLE Lindquester,G.J., O'Brian,J.J., Anton,E.D., Greenmoyer,C.A., Pellett,P.E. and Dambaugh,T.R.
JOURNAL Genetic content of a 20.9 kb segment of human herpesvirus 6B strain Z29 spanning the homologs of human herpesvirus 6A genes U40-57 and containing the origin of DNA replication
Arch. Virol. 142 (1), 193-204 (1997)
MEDLINE 97300856
AUTHORS 5 (bases 106012 to 126166)
TITLE Lindquester,G.J., Greenmoyer,C.A., Anton,E.D., O'Brian,J.J., Pellett,P.E. and Dambaugh,T.R.
JOURNAL Comparison of a 20 kb region of human herpesvirus 6B with other human beta herpesviruses reveals conserved replication genes and adjacent divergent open reading frames
Arch. Virol. 142 (1), 193-204 (1997)
MEDLINE 97300864
AUTHORS 6 (bases 1 to 162114)
TITLE Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S., Inoue,N. and Pellett,P.E.
JOURNAL Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A
J. Virol. 73 (10), 8040-8052 (1999)
MEDLINE 99412318
AUTHORS 7 (bases 19696 to 22886)
TITLE Pellett,P.E.
JOURNAL Direct Submission
Submitted (26-JUN-1993) Herpesvirus Section, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta, GA 30333, USA
8 (bases 64106 to 84963; 106012 to 126166)
JOURNAL Direct Submission
Submitted (02-OCT-1993) Herpesvirus Section, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta, GA 30333, USA
9 (bases 1 to 162114)
JOURNAL Pellett,P.E., Dominguez,G., Dambaugh,T.R., Stamey,F.R., Dewhurst,S. and Inoue,N.
TITLE Direct Submission
Submitted (10-JUN-1999) Herpesvirus Section, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G18, Atlanta, GA 30333, USA
REMARK Sequence updated by submitter
COMMENT On Aug 16, 1999 this sequence version replaced gi:405170 gi:405156 gi:305397.
FEATURES
source location/Qualifiers
1..162114 /organism="Human herpesvirus 6B"
/strain="Z29"
/db_xref="taxon:32604"
/note="variant B"
1..8793 /rpt_type=direct
59..353 /rpt_type=tandem
/note="heterogeneous telomeric unit"
/rpt_unit="caryc
583..3988
gene /gene="DR1"
/note="US22 gene family"
join(583..841,955..2975)
/gene="DR1"
/codon_start=1
/product="DR1"
/protein_id="AAD49614.1"
/db_xref="GI:5733511"

/translation="MPLTARAGHTLRRLPLSHYWLILGRLSHRYSLRLKRLGLRL
PLPPEQECIHLHPKPYKRLRLPCLTLPRLHLOGPADSLIMFEDPKRPPASDSTLL
PLGLITTSARSMKSEPTKSGFHADPSLSMTLGTSSPWLVLQAFGSLFECDFVQ
RLILSHSVLFLKTGLRQCEALYRAVLWVRPLPSLWTCRDPTAFLEPLTLAR
GLAFAFLWRLHLGSRSELSPVLEWERELVLTDRKGPCHLILSGSSQVSSSD
ACGTWNAATEKAGKEAEKRGQOATDRLASPHLTRGRDCGSLGSEPPSAEDF
ARCPILDELCEGGMPLPFAFLTASPVCLILEGGVLAIDLNDLSIRIADLEL
LRGLSLILSGRLPLRPPSGSEAAKPKYEKEEGRRATTSATTAASRRPTPR
GYNKRGVTTGDXVPSFSAHPSEBQTOGHHRRQSGHGDDRGGRGRHRRDGAHND
ETPEQORGEHDEQTDSGREDAQSEVARDEKGTQGGSGRSGCATQYTGKGE
HGAWSTLPSLVPRDPVWPVPHLLFSPSLPSITVEDEPSARHPCPPAPAEPSKC
SPDPCSPDAPQAVPRLSALVSPSTARVFLSLSSSSSSSSSSSSSSSSPSPL
SP
ASPSAPICIGSRPSAQT"
complement(2723..3325)
/gene="DR3"
/evidence=not_experimental
complement(2723..3325)
/gene="DR3"
/codon_start=1
/product="DR3"
/protein_id="AAD49618.1"
/db_xref="GI:5733515"
/translation="MSRVSCVLRACVCAGLCQVCMGVICGDCQRMWRRCARMGRV
GPRVLGAGARVRSGDARSVSRTCTERAPSAKSPVSPVLVSSSSSSSSSAC
SSRVSP
GCPEDHGFARSQDLMGEMNGDLDEGETGEGDNGEGE"
3022..3398
/gene="B1"
/evidence=not_experimental
3022..3501
/gene="B1"
/codon_start=1
/product="B1"
/protein_id="AAD49617.1"
/db_xref="GI:5733514"
/translation="MQNMKTKTKTKRGRKEGNTPTETERMEPARSRTSAISGLRR
SGSTPLRGPVHAPTWRTASATADSHPLSPPTSPSSGRHITRCAARTRTS
AAELNGVAAVAVAKTKRSETIDRLLSVLPGHGPASLRSHLRANSALRPPDDPR"
3536..3988
/gene="B2"
/evidence=not_experimental
3536..3775
/gene="B2"
/codon_start=1
/product="B2"
/protein_id="AAD49616.1"
/db_xref="GI:5733513"
/translation="MQPMTKKTKTKRSPKSPAPPSLSPPLESRKRVGKSPERLHGN
APPTPPRSP
3912..3917
/note="DR1; B1; B2"
/evidence=not_experimental
3938..3942
/note="DR1; B1; B2"
/evidence=not_experimental
3968..3974
/note="DR1; B1; B2"
/evidence=not_experimental
3983..3988
/note="DR1; B1; B2"
/evidence=not_experimental
5027..7435
/gene="DR6"
/note="US22 gene family; transactivator"
join(5027..5330,6529..7203)
/gene="DR6"
/codon_start=1
/product="DR6"
/protein_id="AAD49615.1"
/db_xref="GI:5733512"
/translation="MTRHTQTGRDGRIRARDGARLAAHARARAREWILLARGPSKL
YGTYSRHRGERIHLWPARYWCLELHPDYPDARSATVWGHRWGMPPIHVRPSYODCA

MD 21331, USA
On Jun 17, 1995 this sequence version replaced gi:662095.
COMMENT
FEATURES
location/Qualifiers
1. 21191
/organism="Human herpesvirus 6"
/strain="01102"
/specific_host="Homo sapiens"
/db_xref="taxon:10368"
/clone="HindIII-D, EcoRI-D, pPCR-3"
/map="109-130 kbp"
1. 1980
/gene="HDFR0"
/gene="HDFR0"
/codon_start=1
/protein_id="AA68464.1"
/product="replication origin-binding protein"
/db_xref="GI:662096"

gene
CDS

gene

CDS

/note="component of helicase-primase complex; similar to
HCMV UL102 Swiss-Prot Accession Number P16827"
/codon_start=1
/product="unknown"
/protein_id="AA68465.1"
/db_xref="GI:662097"

polyA_signal

gene

polyA_signal

gene

polyA_signal

gene

polyA_signal

CDS

/note="similar to HH6 CB2L gene product, GenBank Accession
11477. 12652
/gene="HDFR1"

CDS

gene

CDS

/note="putative helicase; similar to HCMV UL105,
Swiss-Prot Accession Number P16736"
/codon_start=1
/protein_id="AA68468.1"
/db_xref="GI:662100"

polyA_signal

gene

polyA_signal

CDS

Number 114772"
/codon_start=1
/product="unknown"
/protein_id="AA68466.1"
/db_xref="GI:662098"
/translation="MKSLLKLELETSDVENLHVRITLKVIRKICISLARRPLVNI
TVDDGCIWFLDEGTMINGLETKICEKEFGFGIGIITLDESDTLOEILNLPQPK
RLIHMKVDPPEEFLMGLGLEALEENLPLKOSTLHKILAKVLEPALSPITKLEFNC
TLYCTDLPHEFNHSAETLAKVPLMIRLNYEMNTHIEVLELYFNKRSHNFINLSLE
SKQLDSSLOVLTQATFOATFYTSKNDIS"
complement(4630. 6618)
/gene="HDFL1"
/note="similar to HCMV UL104, Swiss-Prot Accession Number
P16735"

/codon_start=1
/product="vition protein"
/protein_id="AA68467.1"
/db_xref="GI:662099"
/translation="MRRASANSLLNSVSGSMRRNOSGRRPCKRLSDNEATLSTINS
ILGEMDLKSNLSLYLPNNNEIDMTPSOIMTFIEMLGHKNFPGGOTIHNAIDBS
AVLKQAVCAQALINSVSIQOIHDEMRHVSFPHNKRSLSPDYFSWMAIAIOY
ILGDISITLFIQSIDNSRTRVDMTCVGIAPRRTTHRNPLNLEQYSSKIL
HDIIVRKNQNTVPVLGLSSVLIIPDFHNKIFRDRNSEQISCEKKNKALAEFYSPY
VIRNRMLATPLAHLSPELKNHSLRHRQCOLMTPIKVLITAKDTVTKRKINDM
IEREKSSDAKSLIKELNLSDSKSIGRDSVEGFIQETPSIIDONKMLNRCOF
RRKSAIDTGRDVRDLFKKOIICMEQIOIOMDEITLKTOMPERKIKDLSHLE
TNNDCDYNHNDLLENLSALNLYQRLPTSVSIDPTRVANSFEOYIPDOY
ADRIKIDLMEMETRTLRKNVNNOSQESITYSNLTLLVPLRLRLNLYNLES
IPEPFELEISGETILAIYESSKLKHVLRVLYVELNQISEVFNLTQHPENSBPIDS
NIPSPENLEILKIKRINRLRIQHLTRBNYPRKGD"

6413. 9409
/gene="HDFR2"
6413. 8887

complement(9325. 9330)
/gene="EDLF5"

/evidence-not_experimental
complement(9325. 10351)
/gene="EDLF5"

9404. 9409
/gene="HDFR2"

/evidence-not_experimental
complement(10022. 10351)
/gene="EDLF5"

/codon_start=1
/product="unknown"

/protein_id="AA68469.1"

/translation="MNLALFYLCROLLAAVUIIEPPRLCGTTLAFYSYHPLSTCH
IDSFVTSRTSNTLFLCHLYCHORSHOKSHOKTKTGYLKLFSNTIENIKNNFNSSP
YTLSTTA"

11477. 12652
/gene="EDRF1"

CDS

11477..12511
/gene="EDR21"
/note="putative phosphoprotein; similar to HH6 CB7R,
GenBank Accession Number U14772 and HCMV UL112, Swiss-Prot
Accession Number P16768"
/codon_start=1
/protein_id="AA68470.1"
/db_xref="GI:662102"
/translation="MYAEKAYGSFDDMEYVQOIVTESIQKRLHFGSGCLIEFLG
SGICEIFCGWISMTGWSLTVSDIWDGKTYTNARGFLLCSKSTISQ
NTEGRDLITVSOENNKLIQITFVLTIVKEHDIRNLGDPKIEKFEKCALDRKH
DDEHRRSGKQEKREKVEDTKKDDDRKQERKRNDEKQDPKRESDGKYCIIV
VNFYSCVTVDVFLNSICFVAISLPEKROKHYHDMKRNLEEOSHEDGITLTSTTV
NGAVEGALPDCISIDNHEDQDHDELDRKRAYAQVGVSGSPKPRSSGSLCVSKADLFI
LGDEPRRSIQ"
12483..13180
/gene="EDR2"
/note="EDR2"
/gene="EDR2"
/note="putative phosphoprotein; similar to HCMV UL113,
Swiss-Prot Accession Number P17151"
/codon_start=1
/function="replication"
/protein_id="AA68471.1"
/db_xref="GI:662103"
/translation="GTNREGLSNEDNYGNFRLNKSLEQLRALVASSGDIYERSLKL
KECLDNVKNLKNKCAVDTPSKSLSKTHIEPKQIVFSDCVRPVPCIKRPFIDV

gene
CDS

Query Match 3.7%; Score 83.6; DB 72; Length 21191;
Best Local Similarity 53.3%; Pred. No. 0.0062;
Matches 176; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1608 cccagcagagacctgcagagcagagccagagccatcgagagccagagccagagccccc 1667
DB 19491 cccagcagagacctgcagagcagagccatcgagagccagagccagagccccc 19432
QY 1668 cccgagcagagacctgcagagcagagccagagccatcgagagccagagccccc 1727
DB 19431 cccgagcagagacctgcagagcagagccatcgagagccagagccagagccccc 19372
QY 1728 cccgagcagagacctgcagagcagagccagagccatcgagagccagagccccc 1787
DB 19371 cccgagcagagacctgcagagcagagccatcgagagccagagccagagccccc 19312
QY 1788 cccgagcagagacctgcagagcagagccagagccatcgagagccagagccccc 1847
DB 19311 cccgagcagagacctgcagagcagagccatcgagagccagagccagagccccc 19252
QY 1848 cccgagcagagacctgcagagcagagccagagccatcgagagccagagccccc 1907
DB 19251 cccgagcagagacctgcagagcagagccatcgagagccagagccagagccccc 19192
QY 1908 cccgagcagagacctgcagagcagagccagagccatcgagagccagagccccc 1937
DB 19191 cccgagcagagacctgcagagcagagccatcgagagccagagccagagccccc 19162

Search completed: October 12, 2000, 18:53:59
Job time: 38215 sec